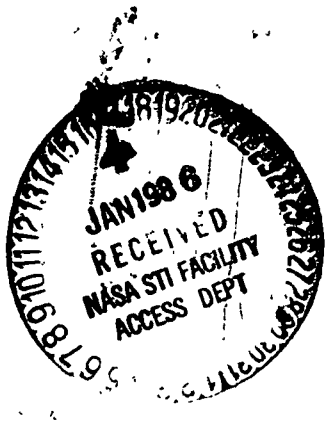
NASA Contractor Report 174797

Robust Detection/Isolation/Accommodation for Sensor Failures

Jerold L. Weiss, Krishna R. Pattipati, Alan S. Willsky, John S. Eterno, and John T. Crawford

Alphatech, Inc.
Burlington, Massachusetts

September 1985



Prepared for
Lewis Research Center
Under Contract NAS 3-24078

NASA
National Aeronautics and Space Administration

CONTENTS

FIGURES

TABLES

SECTION 1

SUMMARY

This report presents the results of a one year study performed by ALPHATECH, Inc. to develop a design methodology for the robust detection, isolation and accommodation (DIA) of sensor failures in jet engine control systems. This study was funded by the NASA Lewis Research Center due to concerns about robustness problems encountered with previous [3] jet engine sensor DIA algorithms. The purpose of this project was to address the fundamental issues of robustness and redundancy in dynamic systems in a quantitative way and develop a framework for designing DIA systems which explicitly tolerate unavoidable modeling errors. The scope of this project during this year has been limited to linear systems, although extensions to specific problems encountered in nonlinear engine models have been outlined. The results of this project as summarized at the end of this section, are believed to represent substantial contributions to the state of the art in failure detection and jet engine control.

This project was organized into the three separate tasks described below.

Task 1. Conduct basic research to develop a methodology for designing a system to detect, and distinguish among, a set of failure modes in the presence of model uncertainties;

Task 2. Demonstrate the design methodology via application to an F-100 engine sensor FDIA problem at a single operating point;

Task 3. Evaluate alternative FDIA approaches proposed for the F-100 engine problem, and, in particular, the FDIA algorithm of NAS 3-22481 [3].

In recent years a wide variety of techniques has been proposed for the detection, isolation and accommodation of failures in dynamic systems (see, for example the surveys in [1],[2] and the numerous methods discussed in [3]-[10]). Some of these methods have been developed starting from general, abstract dynamic models, while others have been produced in the context of particular applications. While the general methods provide the basis for what in principle should be a widely applicable failure detection methodology, their very generality often tends to obscure the important concepts that must be considered in the design of practical and reliable failure detection systems. Conversely, the application-specific methods, which do address these basic concepts, typically offer little insight into how to generalize their results.

As a result, there has not been a satisfactory general design methodology for robust failure detection algorithms. In particular, the general approaches to failure detection described in [1]-[3] take as their starting point mathematical models of both the system under consideration and of the types of failures that may occur. However, if one attempts (as, for example, was done in [3]) to use one of these approaches in a top-down or "canned" manner in which one generates the requisite overall models and then essentially plugs them into the approach chosen, the likely result will be a failure detection algorithm that does not work satisfactorily. The general reason for this is accurately stated in the request for proposals which resulted in this study: "A fundamental limitation of the performance of this (referring to our reference [3]) and all similar analytically redundant schemes is the adequacy of the model used to establish the reference upon which the detection/isolation decision is based and upon which successful

accommodation may depend." This statement raises key questions. How does one determine what aspects of a model are most important in the context of failure detection and isolation? How does one design DIA systems based on such knowledge? How does one measure performance in a way which not only is meaningful but also is analytical enough to provide a useful tool for comparing approaches, pinpointing critical weaknesses and suggesting alternatives for improving performance?

The philosophy of this program extends from the work reported in [4] and [5]. From these studies, three key concepts have been identified. First, all failure detection, isolation, and accommodation (FDIA) methods are based, either implicitly or explicitly, on the use of redundancy relations among the measured system variables. Consequently, the robustness of the FDIA process depends on the reliability of such redundancy relations, given the inevitable presence of model uncertainties and noise.

Secondly, the presence of redundancy does not guarantee that all failures of interest can be detected and distinguished. Clearly, only observable failures are detectable and distinguishability of failure modes is only possible when the effects of each failure are distinct.

Finally, information about the presence or absence of particular failures tends to accumulate over time. Failure decisions which make use of multiple observations of the measured system variables will have lower error probabilities than "single shot" decisions if the proper use of failure information is made.

These concepts have been examined in detail and have resulted in several interesting solutions to the tasks described above. Our design methodology (Task 1) consists of a set of analytical techniques for evaluating various

FDIA designs and a variety of procedures for optimizing and modifying individual components of that design. The evaluation process is based on the notion of probabilistic distance metrics. By including modeling error in the stochastic description of system behavior, we can then interpret these metrics in terms of realizable FDIA performance characteristics. These metrics are also used to optimize the robustness of redundancy relations for a particular range of modeling errors. Since failure coverage and distinguishability is of concern, particular attention is paid to those metrics which relate to detectability and distinguishability of failure hypotheses.

The application of this design methodology to the F-100 engine (Task 2) has resulted in an interesting structure for practical FDIA schemes. This structure consists of two levels and avoids many of the computational problems associated with the so-called "optimal" methods, [1], [2]. At the top level, a monitoring system examines signals for the possible presence of any failure. A separate test for each failure is used to minimize the decision delay following a failure. Alarms are raised at this level and trigger sequential testing procedures which: 1) compare all pairs of failure hypotheses which are potentially ambiguous following a given set of trigger alarms, and 2) verify that the alarm was not a false trigger. The hypothesis tests make use of the residual signals which are derived from the optimization problems in our robust FDIA design methodology. The "parameters" of these tests are determined through a trade-off analysis which makes use of the metrics we have developed. Final failure decisions are then made on the basis of these individual test results.

Finally, in Task 3, we demonstrate the generality of the distance metric evaluation tool by defining methods of evaluation for the FDIA scheme of reference [3]. These schemes are then applied to a similar system and some potential methods of improvement are suggested. These improvements attempt to maintain much of the structure of the original algorithm including the use of Kalman filters for generating residuals. We attempt to identify those parts of the system model which most degrade performance and try to remove that segment from the filtering process.

The remainder of this report is organized as follows. Section 2 introduces some of the fundamental concepts and problems of failure detection and isolation. Section 3 provides an overview of an approach to FDI design which addresses the fundamental problems of Section 2 and requires analytical tools such as those described in Section 4 and Appendix A. In Section 5, we discuss several specific analytical FDI design procedures which make use of the results in Section 4. An example of a complete sensor FDI system for the F-100 engine at a single operating point is also provided and simulation results are presented. We also apply our results to the evaluation of a Kalman filter based FDI algorithm. Finally, Section 6 summarizes the contributions of this project and provides recommendations for future work. Appendix B provides an extension of our analytic results to the robust accommodation problem and Appendix C documents the design and simulation software which was generated for this project.

SECTION 2

INTRODUCTION, MOTIVATION AND BACKGROUND

In order to motivate the analytical results discussed in Section 4 and Appendix A, we present below several pertinent questions and some discussion of each. These questions form a logical sequence, starting from what at first may seem to be the simplest and most transparent query:

- What is failure detection, isolation, and accommodation (FDIA)?

Obviously, FDIA deals with the problem of detecting deviations from normal behavior in a specified components (sensor or effector), isolating the particular component which has "failed", and initiating the appropriate adjustment to minimize the effect of the failure. The key point in this sentence is that in order to detect, isolate, and accommodate deviations, one requires a specification of what "normal behavior" is and of what a "deviation" is (i.e., models of the system and of the size and perhaps the nature of deviations to be detected). Furthermore, for each type of anomaly to be detected, the model must provide sufficient redundant information to allow one to detect the anomaly and to distinguish this anomaly from others. For example, in a triplex sensor system, in which there are three identical sensors of each type, one can perform voting by examining each triple to determine if they are consistent (i.e., normal). Here the model information used is that the three sensors measure the identical quantity, and the model of a deviation can be specified in several ways, such as in terms of

manufacturer's instrument specifications. As a second example, consider a relatively simple and often-used check, in which successive samples of the output of a particular sensor are examined to determine if there is an obvious inconsistency. Here the model information used is a crude measure of the bandwidth of the variable being sensed. Finally, consider a simple system involving linear motion and in which one has a velocity sensor and an accelerometer. Here the kinematic model $\dot{v} = a$ provides a mechanism for obtaining one redundant relationship between these sensors.

In the terminology used by Chow and Willsky [4], [7], the three examples just described are illustrations of direct (or hardware) redundancy, temporal (or self-test) redundancy, and analytic (or functional) redundancy, respectively. While there are clear differences among them, it is their similarities--in terms of being based on models and, more explicitly, on redundancy imbedded in those models--that we wish to stress. This permits us to construct a unified framework in which to examine and compose different approaches to failure detection and their robustness properties.

- What does an FDIA algorithm do?

Here again we follow Chow and Willsky. Roughly speaking, all failure detection systems can be described in terms of the conceptual block diagram of Fig. 2-1. There are three basic parts of the failure detection process. The first part, termed "residual generation" in the figure, uses the model that has been specified to generate "residual" signals which nominally should be near zero and which will deviate from zero in characteristic ways when particular failures occur. The way in which residuals are generated differs markedly from method to method. For example, in a triplex system, if $y_1(k)$,

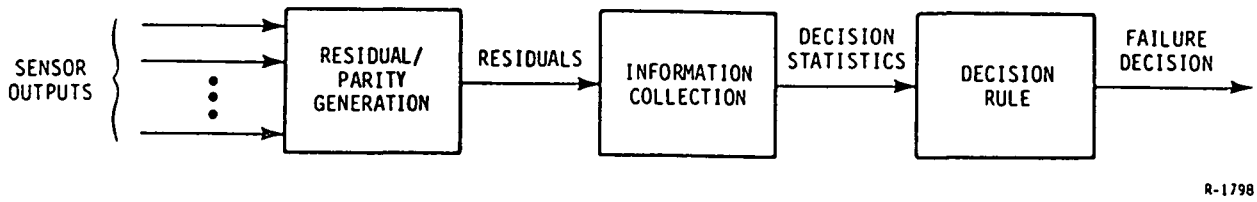


Figure 2-1. Three Stage Structure of the FDI Process

$y_2(k)$, and $y_3(k)$ denote the outputs of three identical sensors, then $r_1(k) = y_1(k)-y_2(k)$ and $r_2(k) = y_2(k)-y_3(k)$ can be thought of as the residuals used in a voting system. In more complex methods, such as the Generalized Likelihood Ratio (GLR) method [1], [2], [7], [9], Kalman filters are used to generate the residuals, while in other advanced methods including the detection filter approach of Beard [10] and Jones [18], Kalman-like filters are designed, but with gains chosen in particular ways so as to make particular failures more readily apparent.

The remainder of the FDIA system is the decision mechanism, which consists of the information collection and decision rule stages. This mechanism involves the examination of the residuals to detect, isolate, and accommodate failures. Depending upon the residual generation procedure, this process may take on quite different forms. In the above triplex system example, a simple rule of the form

$r_1(k)$ small, $r_2(k)$ small	No failure
$r_1(k)$ small, $r_2(k)$ large	y_3 failed
$r_1(k)$ large, $r_2(k)$ small	y_1 failed
$r_1(k)$ large, $r_2(k)$ large	y_2 failed

can be used. Similarly, in the detection filter approach, the vector of residuals is constructed in a way so that relatively simple rules somewhat similar to voting can be used. For the Kalman filter-based methods, such as GLR, the effect of particular failures on the residuals can be calculated, but in general the nature of these effects is not as simple as in the other methods just discussed. Rather, what is required in the information, collection and decision rule stages is a very indirect detection test such as the WSSR method used in [3], which must be followed or replaced by a more complex isolation process such as the GLR procedure of correlating the residuals with failure signatures.

It is here that we begin to see an important distinction. The voting and detection filter approaches involve relatively simple decision mechanisms. This is because they and several other approaches to FDIA (such as the "parameter synthesis" approach in [3]) make explicit use of specific redundant relationships in the residual generation procedure. On the other hand, GLR and other Kalman filter-based methods make only implicit use of such relationships in terms, for example, of the resulting failure signatures. However, the advantage of an implicit approach over the explicit methods is that a Kalman filter-based approach represents a statistically optimal method for extracting and using information embedded in the residuals.

- Why is robustness an issue?

The answer to this question is obvious. As we have argued, all failure detection methods are based on the utilization--explicitly or implicitly--of dynamic models and, more directly on the redundancy relationships such models imply. If the model is in error, the redundancy relationships will also be in

error, and consequently, the behavior of residuals will deviate from their ideal characteristics. For example, in a triplex system, it will never be the case that the three sensors are absolutely identical. As an example consider a problem in which one has three accelerometers or gyros. In this case, scale factor errors and misalignment angles, to mention only two possibilities, introduce errors into the assumption that the instruments are measuring the same variable. In the case of highly complex nonlinear systems such as a jet engine, one uses a linear model to describe behavior at a single operating point. Thus, clearly, errors will exist due to the linearization error as well as the uncertainty in predicting the operating point. Furthermore, these errors vary over the operational envelope of the system as well as during transient operation.

Intuitively, a residual generation procedure attempts to remove the predictable part of sensor outputs and to produce signals whose behavior under ideal conditions is unaffected by the value of the variables being sensed. For example, the signal $r(k) = y_1(k)-y_2(k)$, where y_1 and y_2 are identical sensors, should only deviate from zero due to sensor noise or to a failure in one of the two sensors. However, if there is a scale factor difference between the two, there is now another error source whose magnitude is modulated by the variable being sensed. In the case of a Kalman filter, the vector of residuals r(k) is ideally a white noise sequence uncorrelated with previous measurements. However, if there are any model uncertainties r(k) will in general have nonzero mean and a relatively complex correlation structure.

The preceding paragraph suggests one important point and leads directly to another. The first point is an apparent advantage of techniques such as

voting and parameter synthesis, which make explicit use of redundancy relations. In such a case it is relatively straightforward to deduce the effect of modeling errors on the residuals, and consequently one can use this information to assess the relative merits of different redundancy relationships and can determine how to design decision rules based on those that are deemed to be "robust enough." On the other hand, methods such as those based on Kalman filters (and to a lesser extent detection filters) make only implicit use of redundancy relations, and thus it is apparently far more difficult to assess how uncertainties in these relations affect performance. This is precisely where one finds the difficulty with "top-down" approaches to FDIA algorithm design such as were used in [3]. In particular, in a given system one generally has several sources of redundancy which are of different quality or certainty. A top-down approach using an implicit method (such as GLR, WSSR, or the bank of observers approach) and a full system model in effect mixes together redundancy relations that are accurate with ones that are quite uncertain, resulting in either severe sensitivity problems or severe limitations of detectable failures. A more sensible approach is the separate use of these relations so that one can take optimum advantage of each. Thus, one is led to the concept of identifying the most reliable redundancy relations and designing optimum algorithms based on these. These ideas form the basis of our work reported here and in [12] and [13].

SECTION 3

OVERVIEW OF THE DESIGN APPROACH

In this section, we present a brief overview of the major concepts behind our approach to designing robust FDI algorithms.

The first part of our design methodology is to identify those portions of the system dynamics (called parity relations) that are known with the most certainty, as the use of residuals based on these relations will be of great value in minimizing false alarms. Thus the first problem we wish to address is that of defining some "robustness metric" that quantifies how close to zero each residual is under normal conditions given the presence of model errors and noise. We can then develop a rank ordering of parity relations in terms of robustness and address the next problem which is coverage.

Although solution of the first problem, in principle, provides a set of robust parity relations, there is no guarantee that the failure modes which must be identified are distinguishable from normal operation (i.e., that all failure modes are covered). Thus the second problem is to define a metric to assess the ability of the relations identified in the first step to detect a specified set of failure modes. Each failure mode is also modeled with an allowance for errors so that the effect of uncertainty associated with the failure model will be minimized. As a result, the initial set of relations may be augmented and residuals which are useful in detecting particular failure modes not well-covered by the initial group are generated. Let us note

two aspects of this problem. The first is that it requires a second metric that measures the ability of a particular relation to distinguish between normal conditions and a particular failure mode (i.e., to give decidely larger values under the particular failure than under normal conditions) given the presence of model error and noise. Again this metric can be used to rank-order relations with respect to their usefulness for detecting particular failures. The second point is that the relations added at this stage are less reliable than the first ones obtained, in that they may have larger values when no failure has occurred. They may be needed, however, to achieve the desired coverage (i.e., to achieve a specified probability of detection for all failures). Note also that the metric used should provide a guide to the minimum magnitude of a particular failure that can be reliably detected (i.e., to achieve a specified probability of false alarm).

The final step in the design is concerned with the problem of distinguishability: given that a failure has occurred, can we determine which failure mode has occurred. Here again we need a metric that measures the ability of a parity relation to distinguish a particular failure mode from an alternative set of possible events corresponding to one or more of the other failure modes. Note again that if additional relations are needed, they will be inherently less reliable under normal conditions. However, at this stage we can, if necessary, avoid the impact such relations would have on false alarm rate by using a two-level structure. Specifically, the relations determined in the first two stages are sufficient for detection of all failures, but may not be able to isolate all of them. Thus, one could use these relations to detect and trigger the use of additional relations for isolation only. Fig. 3-1 gives a complete picture of the FDI system design process.

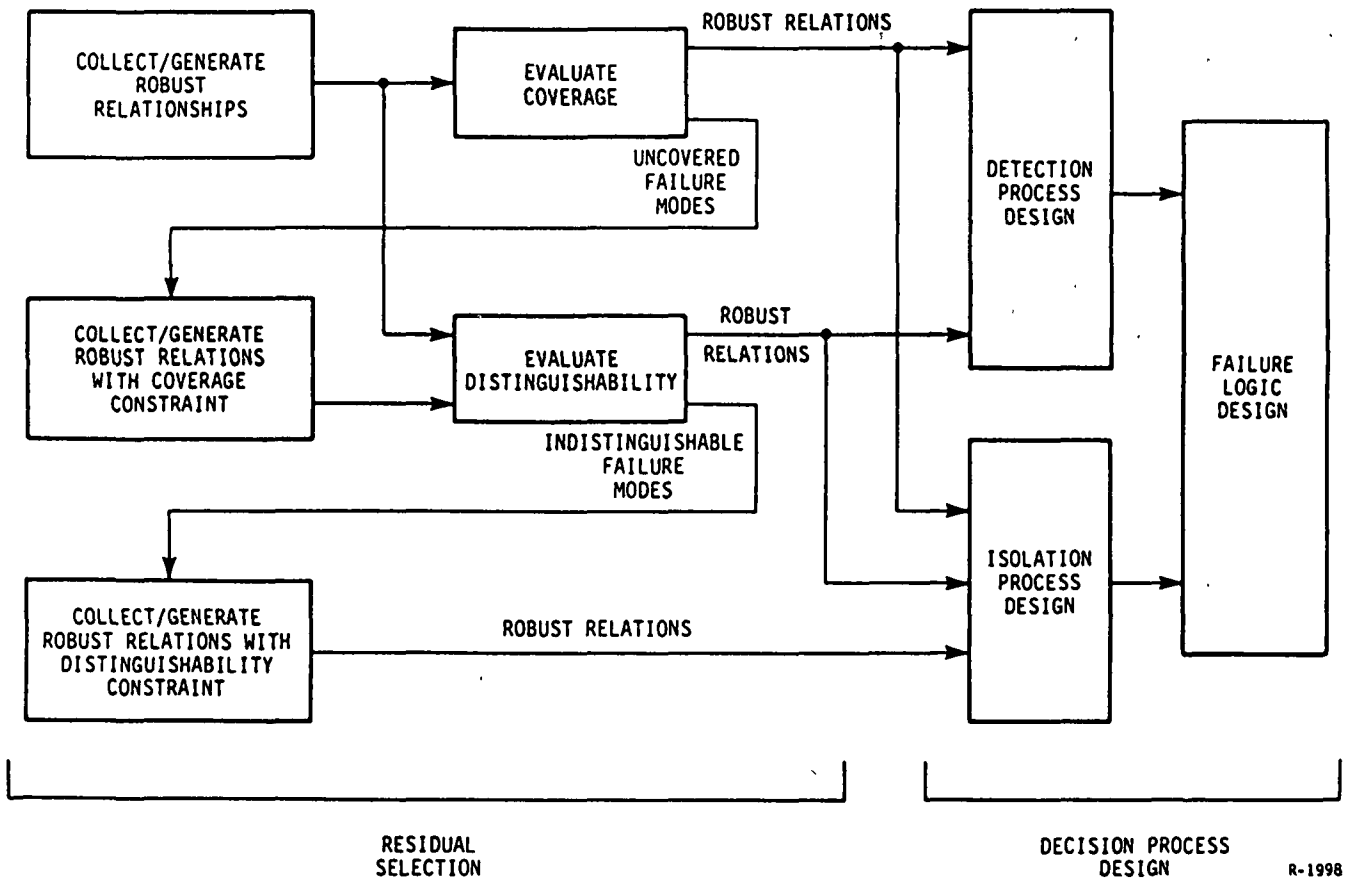


Figure 3-1. Overview of the FDI System Design Process

For the most part we have discussed the residual generation phase of the FDI system. However, the concept of metric-based evaluations with a probabilistic description of model error is also of value in designing the information collection phase, i.e., in determining the data length required to achieve desired performance levels and in specifying the details of how successive residuals are accumulated. These issues are discussed in more detail in Section 5.

In the next section, we describe the metrics we have considered and indicate how they are used in the design of robust FDI systems, and the evaluation of alternative FDI schemes.

SECTION 4

ANALYTICAL RESULTS

In this section we develop the analytical basis for the design and analysis of FDI systems. We start with a definition of redundancy and pose several optimization problems which, when solved, provide relationships among measured variables which can be used in the FDI process. These optimization problems are then interpreted as special cases of statistical discrimination. Various distinguishability "metrics" are discussed which provide the analytical basis for FDI performance evaluation. These metrics also provide an alternative mechanism for generating redundancy relationships which are "optimal" in the sense of the performance metric which is used.

PROBLEM FORMULATION

Consider a linear, discrete time, time-invariant dynamic system with uncertainty characterized by a finite set of system parameters viz.,

$$x(k+1) = A_\ell\, x(k) + B_\ell\, u(k) + E_\ell\, w_\ell(k) \tag{4-1}$$

$$y(k) = C_\ell\, x(k) + D_\ell\, u(k) + v_\ell(k) \tag{4-2}$$

where

$x(k)$ = NS - dimensional state vector at time k,

$y(k)$ = NO - dimensional measurement vector,

$u(k)$ = NC - dimensional measurable control vector,

$w(k)$ and $v(k)$ are process and measurement noises respectively and are assumed to be zero-mean, white, and Gaussian with covariance matrices Q_ℓ and R_ℓ

respectively, and where $\ell = 1,2,\ldots,L$ with the a priori probability of the ℓth model being correct denoted by p_ℓ. (Note that it is indeed possible to formulate this problem using a continuum of parameter variations. Similar results may then be derived).

A redundancy relationship is now defined as a linear combination of measurements and controls over a finite window of observation. Specifically, if we let $Y_p^T(k) = \left(y^T(k), y^T(k+1),\ldots y^T(k+p)\right)$, and $U_p^T(k) = \left(u^T(k), u^T(k+1),\ldots u^T(k+p)\right)$, then redundancy relations take the form

$$\underline{\nu}(k) = W^T \begin{bmatrix} Y_p(k) \\ U_p(k) \end{bmatrix} = W_y^T\, Y_p(k) + W_u^T U_p(k) \tag{4-3}$$

where $\nu(k)$ is the t-dimensional residual vector which, under ideal circumstances (no noise or modeling error) is identically zero. The matrix W is sometimes referred to as the parity check matrix. Next, we can expand $Y_p(k)$ in terms of the parameters of the ℓth system model as

$$\begin{aligned} Y_p(k) = &\begin{bmatrix} C_\ell \\ C_\ell A_\ell \\ \vdots \\ C_\ell A_\ell^p \end{bmatrix} x(k) + \begin{bmatrix} D_\ell & 0 & \cdots & 0 \\ C_\ell B_\ell & D_\ell & & \\ \vdots & & & \vdots \\ C_\ell A_\ell^{p-1} B_\ell & \cdots & C_\ell B_\ell & D_\ell \end{bmatrix} U_p(k) \\ &+ \begin{bmatrix} 0 & \cdots & 0 \\ C_\ell E_\ell & 0 & \vdots \\ \vdots & & \\ C_\ell A_\ell^{p-1} E_\ell & \cdots & C_\ell E_\ell \quad 0 \end{bmatrix} \begin{bmatrix} w(k) \\ \vdots \\ w(k+p) \end{bmatrix} \\ &+ \begin{bmatrix} v(k) \\ \vdots \\ v(k+p) \end{bmatrix} \end{aligned} \tag{4-4}$$

or

$$Y_p(k) = M_{p\ell}\, x(k) + N_{p\ell}\, U_p(k) + \Pi_{p\ell}\, W_p(k) + V_p(k) \tag{4-5}$$

Thus under the ℓth model hypothesis Eq. 4-3 can now be written as,

$$\underline{\nu}(k) = W^T \left(\begin{bmatrix} M_{p\ell} & N_{p\ell} \\ 0 & I \end{bmatrix} \begin{bmatrix} x(k) \\ U_p(k) \end{bmatrix} + \begin{bmatrix} \Pi_{p\ell} \\ 0 \end{bmatrix} W_p(k) + \begin{bmatrix} V_p(k) \\ 0 \end{bmatrix} \right) \tag{4-6}$$

Now, consider the simplest case when no modeling error or noise is present. We can make $\underline{\nu}(k)$ identically zero by choosing W as an orthogonal basis for the left null-space of the matrix,

$$\overline{M}_p \triangleq \begin{bmatrix} M_p & N_p \\ 0 & I \end{bmatrix} \tag{4-7}$$

That is, we find all the vectors for which $w^T\overline{M}_p = 0$ and form the parity check matrix using these vectors for its rows.

Comments

1. The number of independent parity checks for any p is NO(p+1)-NS.
2. As discussed in [19], one need only look at values of p=0,...,NS to find all of the independent parity checks.
3. The solution for W can also be obtained by finding the vectors which satisfy $W_y{}^T M_p = 0$, and then solving $W_y{}^T N_p + W_u{}^T = 0$ where $W^T = (W_y{}^T, W_u{}^T)$.

Table 4-1 presents the details of the linear reduced-order F-100 engine model used to demonstrate the analytical results in this project. This model

TABLE 4-1. DETAILS OF F-100 ENGINE MODEL

Model

--- A (NS,NS) ---

8.7763920E-01	9.5448047E-02	-7.6181483E-03	3.2125510E-02
2.3492503E-03	9.3342350E-01	2.9041661E-02	-8.4782876E-02
3.8686348E-04	-1.4044064E-04	9.8545470E-01	1.1503949E-04
3.9211614E-03	-9.3381410E-04	-7.8085437E-04	9.5632940E-01

--- B (NS,NC) ---

4.4787161E-02	6.4114124E-02	-1.5856765E-02	2.9632282E-03	-3.6747955E-02
2.1774249E-02	1.4993310E-02	-1.2106993E-03	-2.1525344E-03	-1.2677516E-02
1.4233845E-03	9.6010260E-04	-7.8124125E-05	-2.6199684E-06	9.1238425E-04
1.7201110E-03	5.1628816E-04	2.1082116E-04	-1.6600619E-05	3.7244798E-03

--- C (NO,NS) ---

1.0000000E+00	0.0000000E+00	0.0000000E+00	0.0000000E+00
-0.0000000E+00	1.0000000E+00	0.0000000E+00	0.0000000E+00
-3.6199179E-02	8.6627400E-01	-1.7538881E-02	-1.4256181E-02
2.3806910E-01	4.1814116E-03	-2.3846535E-02	-2.1576596E-02
-1.3806290E+00	-6.4474080E-01	-2.1066390E-01	2.4134478E-02

--- D (NO,NC) ---

0.0000000E+00	0.0000000E-00	0.0000000E+00	0.0000000E+00	0.0000000E+00
0.0000000E+00	0.0000000E+00	0.0000000E+00	0.0000000E+00	0.0000000E+00
2.2995210E-01	-4.2003830E-01	2.9931962E-02	4.6308138E-03	-7.1547480E-01
1.1445380E-01	-5.3470520E-01	4.3202233E-02	1.8882763E-04	1.2127620E+00
9.3831810E-01	5.8143690E-01	-1.8835608E-02	-3.4454153E-03	3.4962920E-01

Model Uncertainty (for Robust Redundancy)

---DEL A (NS,NS) ---

1.5380000E-02	2.1089999E-02	1.4011107E-03	1.2169647E-03
3.1333333E-03	9.1000004E-03	8.7536051E-04	1.9486381E-04
1.7636009E-03	1.1241082E-03	2.4200000E-03	1.1208000E-03
4.4277371E-03	2.8102705E-03	2.7039999E-03	7.2799991E-03

---DEL B (NS,NC) ---

8.5195694E-03	1.6469174E-03	5.1843788E-04	2.2065216E-04	8.6416323E-03
2.4867395E-03	4.5617996E-04	1.5556102E-04	8.5836473E-05	2.6413030E-03
8.5665914E-04	1.7833420E-04	5.5528384E-05	1.6012165E-05	8.5322355E-04
2.1998493E-03	4.4917196E-04	1.3978184E-04	3.9906063E-05	2.1532946E-03

--- DEL C (NO,NS) ---

0.0000000E+00	0.0000000E+00	0.0000000E+00	0.0000000E+00
0.0000000E+00	0.0000000E+00	0.0000000E+00	0.0000000E+00
6.8665206E-02	1.2315690E-01	3.7289455E-03	4.6490412E-03
3.6485530E-02	1.9306340E+00	1.9171619E-03	3.1758063E-03
1.1721950E-01	9.3263514E-02	9.9560004E-03	1.3030000E-02

--- DEL D (NO,NC) ---

0.0000000E+00	0.0000000E+00	0.0000000E+00	0.0000000E+00	0.0000000E+00
0.0000000E+00	0.0000000E+00	0.0000000E+00	0.0000000E+00	0.0000000E+00
1.1518222E-02	3.2224100E-02	6.0293591E-03	3.0286349E-03	1.6963300E-01
6.5602008E-03	2.4830708E-02	3.7279800E-03	2.7165335E-04	4.1060939E-02
6.2302962E-02	4.0404614E-02	8.7860543E-03	3.3902435E-03	2.8485480E-01

was developed in [11] along with an estimate of the uncertainty of each matrix element. This uncertainty information will be used to generate subsequent results. The nominal operating point corresponds to the maximum power lever angle (PLA = 83 degrees) and the ambient conditions associated with flight at the sea-level-static Mach (M) and altitude (h) condition; M = 0.6 and h = 10,000 feet. The sampling interval is Δt = .02 seconds. The states, outputs and controls represent perturbations from the nominal values. The engine variables (defined in [11]) corresponding to these perturbations are,

$$x = [N1,\ N2,\ T_{t4},\ T_{t4.5}]$$

$$u = [WF,\ AJ,\ FGV,\ SVA,\ BLC]$$

$$y = [N1,\ N2,\ P_{t4},\ P_{t6},\ FTIT]$$

Parity checks, W, generated from the left null space of $\bar{M}_p$ (with A, B, C, D defined in Table 4-1) for p=0 and 1 are shown in Table 4-2. The first line in each parity check corresponds to the coefficients which multiply the perturbations y(k-p) and u(k-p), the next line multiplies y(k-p+1) and u(k-p+1) and so on up to y(k) and u(k).

ROBUST REDUNDANCY

The above results are easily extended for the case of model uncertainty. Since the general residuals formed by $W^T M_{p\ell}$ cannot be zero for every model we pose the problem,

$$\min J = E_\ell \{ ||W^T M_{p\ell}||^2 \} \qquad (4\text{-}8)$$

$$\text{subject to } W^T W = I$$

TABLE 4-2. PARITY CHECK FOR F-100 ENGINE

LEFT NULL SPACE OF $\bar{M}_p$

P	Scaled Sensor Coefficients					Scaled Control Coefficients				
	N1	N2	PT4	PT6	FTIT	WF	AJ	FGV	SVA	BLC
1	0.0927	0.1180	-0.1272	-0.1272	-0.3316	0.0352	0.0352	0.0000	0.0008	0.3259
	-0.0688	-0.2678	0.3590	-0.0089	0.3345	-0.4048	-0.0543	-0.0039	-0.0005	0.1472
1	0.3068	0.0997	-0.2785	-0.4249	0.2423	-0.1275	-0.4983	0.0338	0.0025	0.2402
	0.0428	0.2579	-0.0216	-0.2374	-0.1069	0.1324	-0.0739	0.0089	-0.0002	0.3099
0	-0.0863	0.4048	-0.4662	0.3119	0.0035	0.0682	-0.0311	0.0005	0.0021	-0.7131
1	0.4017	0.1319	-0.0024	-0.0017	-0.1340	0.1600	0.1213	-0.0132	0.0012	0.0205
	-0.4320	0.3419	-0.4505	-0.0248	0.1783	-0.0609	-0.3061	0.0179	0.0027	-0.3546
1	0.4333	0.0721	0.2728	0.0732	0.1347	-0.1800	0.0971	-0.0136	-0.0004	0.0460
	-0.5512	-0.4604	0.1453	-0.0196	-0.1964	0.1532	0.1647	-0.0072	-0.0013	0.1964
1	0.0359	0.1126	-0.1272	0.3139	-0.0681	0.0685	-0.1697	-0.0149	0.0009	-0.4570
	0.0145	0.0362	0.0055	-0.4717	0.1017	-0.0427	-0.3090	0.0221	0.0004	0.5405
1	-0.0303	-0.4924	0.5286	-0.0539	-0.0447	-0.0736	0.2191	-0.0143	-0.0026	0.4593
	0.1221	0.1359	-0.1139	-0.2586	0.0465	0.0122	-0.2131	0.0155	0.0007	0.2159

The constraint in Eq. 4-8 insures that $W \neq 0$. The solution for W now involves an eigenvalue decomposition of the matrix $C_0 = E_\ell \{M_{p\ell} M_{p\ell}^T\}$. That is, each row of W satisfies,

$$w_i^T C_0 = \lambda_i w_i^T \tag{4-9}$$

The optimal value for J is $J^* = \sum_{i=1}^{t} \lambda_i$ so that one easily sees that the t-best redundancy relationships are the t eigenvectors of C_0 corresponding to the t smallest eigenvalues. This approach is referred to as the Robust Redundancy Null Space Approach.

Now, in contrast to the previous formulation which did not include uncertainty, there are no longer a finite number of independent parity checks; p may take on any value. One would expect, however, that W would become ill-conditioned (as p gets large) if only the best parity checks for each p are used. That is, as p grows past the maximum for perfectly known systems (Lou's result [19]), some parity checks may be nearly dependent resulting in a parity check matrix which is nearly singular. The number of parity checks t and the "order" - p are design parameters which need to be chosen in the design of the decision process. Typically, p is chosen only up to NS in keeping with Lou's result [19].

Table 4-3 shows the 10 best parity checks of order p = 0, 1 for the F-100 engine. Each metric, λ_i, in the table can be no larger than 5.62 for p = 0 and 8.60 for p = 1. (The largest eigenvalues of C_0 for each p). A close examination of Table 4-1 points out an important fact discussed in the previous section. That is, a set of redundancy relations does not guarantee coverage of all failure modes of interest. This can be seen as follows.

The sensitivity of each parity check to a sensor bias failure can be derived from the sum of the coefficients which multiply that particular measurement. That is, a bias of size b_i in sensor i results in a bias in the residual ν of size $b_i[e_i^T, e_i^T, \ldots e_i^T]w_y$ where e_i is a unit vector in the "direction" of sensor i, and w_y is the parity check corresponding to ν. Referring to Table 4-3 we may conclude that many of these parity checks are not particularly sensitive to sensor #1 (N1) failures (e.g., for the first parity check the N1-failure-sensitivity is -.5105 + .6851 = .0946. The tenth parity check has N1-failure-sensitivity of -.1049). The usefullness of any

TABLE 4-3. PARITY CHECKS FOR F-100 ENGINE
SMALLEST EIGENVALUES Of Co
(Robust Redundancy - Null Space Approach)

NO CONSTRAINTS

P	METRIC	Scaled Sensor Coefficients					Scaled Control Coefficients				
		N1	N2	PT4	PT6	FT1T	WF	AJ	FGV	SVA	BLC
1	2.604E-04	-0.5905	-0.3240	0.0044	-0.0013	0.0031	-0.0408	-0.0481	0.0109	-0.0014	0.0307
		0.6851	0.2676	0.0108	0.0038	0.0056	-0.0084	0.0033	-0.0003	0.0000	0.0010
1	3.155E-03	0.2435	-0.4618	0.0676	-0.2346	0.0156	-0.0029	-0.0974	0.0064	0.0010	0.3236
		-0.2408	0.5366	-0.0880	0.2476	0.0049	-0.0127	0.0906	-0.0085	0.0004	-0.3603
1	6.576E-03	-0.1797	0.3313	-0.0350	-0.3326	-0.0340	0.0806	-0.1784	0.0180	-0.0019	0.3938
		0.1427	-0.4372	0.0634	0.3587	-0.0105	-0.0444	0.2230	-0.0177	-0.0003	-0.3936
1	2.087E-02	0.0767	-0.0111	-0.5885	-0.1099	-0.0661	0.2130	-0.2693	0.0169	0.0039	-0.2395
		-0.0129	0.1735	0.5089	0.1145	0.1279	-0.2604	0.2043	-0.0157	-0.0017	0.1648
0	2.147E-02	-0.0856	-0.5930	0.6019	-0.0041	-0.0674	-0.0703	0.2972	-0.0172	-0.0029	0.4249
1	2.814E-02	-0.0700	-0.4761	0.3242	-0.0067	-0.0962	0.0239	0.1972	-0.0114	-0.0013	0.2510
		-0.1573	-0.3452	0.4754	0.9921	-0.0700	-0.0394	0.2493	-0.0143	-0.0023	0.3356
0	4.446E-02	0.6522	0.1035	0.2382	0.0015	0.4500	-0.5239	-0.1689	0.0036	0.0006	0.0415
1	5.377E-02	0.2415	0.0120	0.0287	-0.0276	0.5011	-0.5608	-0.3518	0.0211	0.0007	-0.0564
		0.1678	-0.0281	0.1357	0.0309	-0.2622	0.2388	0.2393	-0.0111	-0.0016	0.1247
1	6.272E-02	0.4126	0.0696	0.2580	0.0076	0.1087	-0.1605	0.0741	-0.0097	-0.0004	0.1282
		0.3722	-0.0183	0.1482	-0.0060	0.4616	-0.5238	-0.2241	0.0068	0.0011	-0.0141
0	5.500E-01	-0.1049	0.0356	0.1948	0.3466	0.264	-0.2453	0.5595	-0.0441	-0.0017	-0.6736

parity check for detecting a bias failure in Sensor 1, however, depends on its robustness metric as well as its sensitivity to failures.

ROBUST DETECTION OF SENSOR FAILURES

In the case of sensor failures, a relatively straightforward modification of the robust redundancy problem can be formulated. Since sensor failures show up in fixed directions in the space of measurable quantities we can think of minimizing the metric in Eq. 4-8 subject to the constraint of fixed sensitivity to a particular sensor direction. That is, we want to choose the w which minimizes

$$J^* = \min E_\ell \{ ||w^T \bar{M}_{p\ell}||^2 \} \qquad (4\text{-}10)$$

subject to

$$w^T \begin{bmatrix} e_i \\ \cdot \\ \cdot \\ \cdot \\ e_i \end{bmatrix} \triangleq w^T b_i = K$$

The solution to Eq. 4-10 is obtained by forming the Lagrangian function and taking derivatives with respect to w. The solution is given by

$$w = \lambda \cdot C_0^{-1} b_i \tag{4-11}$$

where $\lambda = K/b_i^T C_0^{-1} b_i$ and the optimal value of the cost function is $J^* = \lambda \cdot K$. This result is intuitively pleasing since it points directly to the tradeoffs which must be made in designing an FDI system, namely, that greater sensitivity to failures (as embodied by large K and relating directly to the probability of detection, P_d) is obtained only at the expense of decreased robustness (as embodied by large values of J^* and relating directly to the probability of false alarm, P_{FA}).

Note that the above solution implies that there is only one parity check for each value of p which makes sense. Table 4-4 shows the result of applying this result (sometimes referred to as the robust detection null space approach) to the F-100 engine model for p = 0,1 and normalizing K so that $w^T w = 1$. The number in the metric column is $w^T C_0 w$ and may be compared directly to the metrics obtained in Table 4-3 to assess relative robustness.

As an alternative to the problem posed in Eq. 4-10, we may consider the optimization of a metric which represents the failure signal to model noise ratio. That is,

$$w = \arg\max J = (w^T b_i)^2 \,/\, E_\ell \{ \, ||w^T M_{p\ell}||^2 \, \} \tag{4-12}$$

Interestingly, the solution is identical to Eq. 4-11 where the scalar λ may take on any value. That is, for additive sensor failures, minimizing the mean square value of the residual with a fixed failure sensitivity is equivalent to maximizing the signal-to-noise ratio.

TABLE 4-4. PARITY CHECKS FOR F-100 ENGINE
(Robust Detection - Null Space Approach)

Sensor Sensitivity Constraint	P	METRIC	SENSITIVITY CONSTRAINTS: Scaled Sensor Coefficients N1	N2	PT4	PT6	FTIT	Scaled Control Coefficients WF	AJ	FCV	SVA	BLC
1	0	4.487E-02	0.6828	0.2472	0.0656	-0.0023	0.4465	-0.4577	-0.2341	0.0080	0.0013	-0.0678
2	0	2.231E-02	0.1357	0.6185	-0.5738	0.0063	0.1052	0.0247	-0.2999	0.0168	0.0028	-0.4097
3	0	2.266E-02	0.0350	-0.5573	0.6469	0.0023	0.0194	-0.1659	0.2625	-0.0164	-0.0027	0.4154
4	0	8.543E-01	-0.0508	0.2525	0.0944	0.5901	0.0682	-0.1552	0.3104	-0.0268	-0.0002	-0.6714
5	0	4.391E-02	0.6411	0.2752	0.0523	0.0044	0.4653	-0.4741	-0.2471	0.0085	0.0014	-0.0895
1	1	3.795E-04	-0.5640	-0.3169	0.0023	0.0021	0.0195	-0.0570	-0.0567	0.0111	-0.0013	0.0211
			0.7021	0.2831	0.0111	0.0001	0.0188	-0.0212	-0.0062	0.0001	0.0000	0.0024
2	1	8.546E-04	0.6319	0.3088	-0.0570	-0.0024	0.0113	0.0382	0.0159	-0.0094	0.0020	-0.0641
			-0.6877	-0.1201	-0.0690	0.0005	0.0117	0.0039	-0.0357	0.0020	0.0003	-0.0487
3	1	6.366E-03	-0.5427	-0.4167	0.1986	0.0044	-0.0042	-0.0735	0.0463	0.0044	-0.0023	0.1495
			0.6170	-0.0048	0.2218	-0.0009	-0.0001	-0.0500	0.0930	-0.0057	-0.0009	0.1462
4	1	4.984E-03	-0.4789	-0.3207	0.0089	-0.1895	0.0010	-0.0125	-0.1374	0.0177	-0.0015	0.2586
			0.5498	0.2829	0.0114	0.2450	0.0063	-0.0363	0.1299	-0.0109	0.0000	-0.2913
5	1	6.190E-03	-0.3127	-0.3425	-0.0089	0.0095	0.1126	-0.1438	-0.1044	0.0110	-0.0003	-0.0225
			0.6909	0.4805	0.0012	-0.0073	0.1152	-0.1138	-0.0706	0.0028	0.0004	-0.0197

STATISTICALLY BASED METRICS FOR ROBUST FDI

Ideally, the parity relations which are of most use are those which would allow us to minimize the probability of making erroneous decisions. However, in most situations, direct minimization of the error probability, so as to determine an optimal set of parity relations, is not possible. This is because an analytic expression for the probability of error is often difficult to come by, and even if it can be found, the expression is too complicated for optimization. Therefore, it is useful to search for criteria that are easier than the error probability to evaluate and optimize but are, in some sense, related to the error probability. Statistical distance or divergence measures between two probability distributions under two hypotheses (normal and failure mode i, or failure mode i and failure mode j) provide such easily compatible criteria.

Reference [12] (Appendix A) identifies (among others) two useful distance measures which we consider here; the J-divergence and the Bhattacharyya distance. The J-divergence between the two probability density functions (pdfs) $p(\nu|H_i)$ and $p(\nu|H_j)$ is defined by,

$$J_{ij} = \int_{\nu} [p(\nu|H_i) - p(\nu|H_j)] \ \ell n \frac{p(\nu|H_i)}{p(\nu|H_j)} \ d\nu \tag{4-13}$$

and the Bhattacharyya distance, B_{ij}, is defined as

$$B_{ij} = -\ell n \int_{\nu} [p(\nu|H_i) \quad p(\nu|H_j)]^{1/2} \ d\nu \tag{4-14}$$

Thus in order to compute and/or optimize Eqs. 4-13 and 4-14 we must have (or approximate) the pdf's under different hypotheses. Following [13], we assume that under the ith hypothesis, and ℓth model, the system is,

$$x(k+1) = A_{\ell}^{i} \ x(k) + E_{\ell}^{i} \ w_{\ell}(k) + d_{\ell}^{i}(k) \tag{4-15}$$

$$y(k) = C_{\ell}^{i} \ x(k) + v_{\ell}^{i}(k) + b_{\ell}^{i}(k) \tag{4-16}$$

where d_{ℓ}^{i} and b_{ℓ}^{i} account for additive failure effects (e.g., bias, drifts). Thus the measurements y(k) (and hence the residuals $\underline{\nu}(k)$) under the ℓ'th model are Gaussian random variables which can be characterized by a steady state mean vector and covariance matrix, and $p(\nu|H_i)$ is a weighted sum of Gaussians, (WSOG). Computation of J_{ij} and B_{ij} in terms of the parameters in Eqs. 4-15 and 4-16, and the parity check matrix, W, is now possible in principle.

However, solutions to Eq. 4-13 and 4-14 for the WSOG distribution are very unwieldy (though computable) and difficult to optimize. A more useful result is obtained if, for each hypothesis, we approximate the WSOG function by a single Gaussian function. The best Gaussian function to choose depends on the goals of the approximation and one can, in general, conceive of optimizing statistical distance criteria in choosing the parameters of the Gaussian approximation. However, one easily computable and commonly used approximation is the Gaussian function which preserves the first two moments of the WSOG distribution. In particular, if we let m_ν^i and C_ν^i represent the mean and covariance (both functions of W) of $p(\nu|H_i)$, we have,

$$m_\nu^i = E_\ell \{\nu|H_i\} = \sum_{\ell=1}^{L} P_\ell \bar{\nu}^{i\ell} \tag{4-17}$$

$$C_\nu^i = \text{cov}\{\nu|H_i\} = \sum_{\ell=1}^{L} P_\ell \{ C_\nu^{i\ell} + (\bar{\nu}^{i\ell} - m_\nu^i)(\bar{\nu}^{i\ell} - m_\nu^i)^T \} \tag{4-18}$$

where the quantities

$$\bar{\nu}^{i\ell} = E(\nu|H_i, \ell\text{th model}) \text{ and}$$

$$C_\nu^{i\ell} = \text{cov}(\nu|H_i, \ell\text{th model})$$

are directly computable from Eq. 4-15 and 4-16 and the parity check matrix. In fact, it is easy to show that $m_\nu^i = W^T \bar{Y}_p^i$ and $C_\nu^i = W^T C_{Y_p}^i W$ where $\bar{Y}_p^i$ and C_y^i are the mean and autocovariance function (ACF) of the window of measurements and are easily computed as in [13].

Using the Gaussian sum approximation embodied in Eq. 4-17 and 4-18 and letting C^i and C^j denote the ACF of Y_p under the ith and jth hypotheses

respectively, we can explicitly evaluate the distance measures in Eqs. 4-13 and 4-14 as:

$$J_{ij} = \frac{1}{2}\,[\bar{Y}_p{}^i - \bar{Y}_p{}^j]^T\, W \left[\left(W^T C^j W\right)^{-1} + \left(W^T C^i W\right)^{-1} \right] W^T\, [\bar{Y}_p{}^i - \bar{Y}_p{}^j]$$

$$+ \frac{1}{2}\,\mathrm{tr}\,\left[\left(W^T C^j W\right)^{-1}\left(W^T C^i W\right)\right.$$

$$\left. + \left(W^T C^i W\right)^{-1}\left(W^T C^j W\right) - 2I_t\right] \qquad (4\text{-}19)$$

$$B_{ij} = \frac{1}{2}\,\ln \det\,\left[\left(W^T C^j W\right)^{0.5}\left(W^T C^i W\right)^{-0.5}\right.$$

$$\left. + \left(W^T C^i W\right)^{0.5}\left(W^T C^j W\right)^{-0.5}\right]$$

$$+ \frac{1}{2}\,[\bar{Y}_p{}^i - \bar{Y}_p{}^j]^T\, W\, \left[W^T\left(C^i + C^j\right) W\right]^{-1} W^T [\bar{Y}_p{}^i - \bar{Y}_p{}^j] \qquad (4\text{-}20)$$

The general optimization approach for both distance measures involves a gradient-type scheme. However, in two special cases, which are of considerable interest in the FDI problem, the optimization provides explicit solutions.

IDENTIFICATION OF SENSOR BIAS FAILURES OF KNOWN MAGNITUDE

In this case only $b_\ell{}^i$ in Eq. 4-16 is different for the hypotheses H_i, $i=0,1,\ldots,NO$. Under H_0, $b_\ell{}^i = 0$. This implies that the ACFs $C^i = C^j = C$, and

the distance measures (Eqs. 4-19, 4-20) are scaled versions of the Mahalonbis distance:

$$M_{ij} = \frac{1}{2} J_{ij} = B_{ij}$$

$$= \frac{1}{2} [\bar{Y}_p{}^i - \bar{Y}_p{}^j]^T W(W^T C W)^{-1} W^T [\bar{Y}_p{}^i - \bar{Y}_p{}^j] \qquad (4\text{-}21)$$

The optimal parity relation, W, which maximizes the criterion in Eq. 4-21 is given by

$$W = C^{-1} [\bar{Y}_p{}^i - \bar{Y}_p{}^j]$$

and the optimal distance measure is

$$M_{ij} = \frac{1}{2} J_{ij} = B_{ij}$$

$$= \frac{1}{2} [\bar{Y}_p{}^i - \bar{Y}_p{}^j]^T C^{-1} [\bar{Y}_p{}^i - \bar{Y}_p{}^j]$$

The parity relation $\nu(k) = W^T Y_p(k)$ described above (i.e., for sensor bias failures of known magnitude) can be thought of as an approximate whitener followed by a correlator, as shown in Fig. 4-1. The approximation stems from the Gaussian sum approximation. Finally, we note that the residual which results in this case is precisely the decision statistic obtained when we consider the problem of distinguishing known signals with (p+1) observations in zero-mean colored Gaussian noise with autocovariance function C and signal means $\bar{Y}_p{}^i$ and $\bar{Y}_p{}^j$.

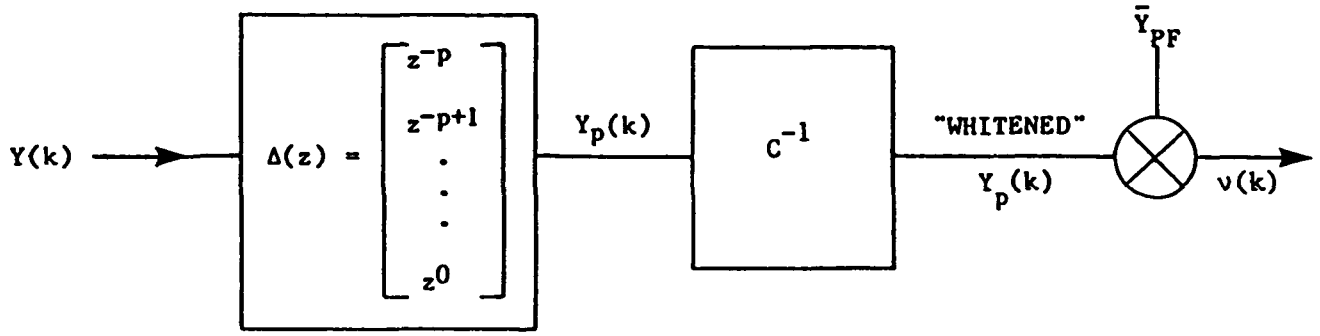


Figure 4-1. Whitener-Correlator Interpretation of Detection Parity Relation

IDENTIFICATION OF NOISE VARIANCE AND SCALE FACTOR FAILURES

This case corresponds to having different covariances under H_i, i=0, 1, ...NO with the additive bias terms $d_\ell^i(k)$ and $b_\ell^i(k)$ zero for all i and ℓ. As a result $\bar{Y}_p{}^i = 0$, but $C^i \neq C^j$. With this simplification, optimization of J_{ij} and B_{ij} yield identical optimal parity relations (Appendix A) which satisfy the eigenvalue equation,

$$C^{i\,-1} C^j \, w = \lambda \, w \tag{4-22}$$

The eigenvectors which make J_{ij} and B_{ij} largest are those which correspond to the eigenvalues for which $\lambda + \lambda^{-1}$ is smallest.

DISCUSSION

The use of statistical discrimination metrics in defining robust parity relations provides us with several interesting results.

1. As discussed above and in Appendix A, the parity checks are functions of the statistical characterization of the measurements (and control inputs if

they are available). This characterization, in the form of means and autocovariance functions, was computed from the uncertain system model of Eqs. 4-15 and 4-16. We can, however, easily of derive parity checks from experimentally determined statistical information as well. Furthermore, insights into the problem of adaptive FDI may be drawn by considering the operation of on-line estimation of the statistics of the measurements.

2. Although we have focussed only on the problem of determining parity check relationships, the metric based evaluation and optimization technique can be applied to the information collection or decision algorithm as well. For example, we might define a window of residuals and maximize the distinguishability of two hypotheses by choosing a linear transformation of the residuals over time. We could also define simpler functional relationships between residuals and decision statistics and optimize the parameters of the relationships using an appropriate metric. Section 5 treats some specific algorithms in this way.

3. In addition to defining parity relations, discrimination metrics such as J_{ij} and B_{ij} are useful in defining performance bounds for any FDI algorithm in the presence of model uncertainty. As discussed in Appendix A and [13], both metrics can be used to place bounds on the performance of the "optimal" decision rule based on the observed data. This allows fundamental performance limits to be determined and used for comparison in the design process. Also, the comparison of various system "configurations" (e.g., various sensor complements including hardware redundancy or various combinations of models) can be accomplished without designing an entire FDI system simply by measuring the distinguishability of various hypotheses with metrics that use the probability density function of the observed quantities.

4. Evaluation of alternative FDI schemes, such as Kalman Filter based algorithms, is possible by applying distinguishability metrics to the sufficient statistics of that algorithm; i.e., using the pdf of these statistics in the metric computations. Section 5 documents an example of the evaluation of Kalman filter based algorithms using metrics applied to the residuals. Because of the invertibility of the Kalman filter, these results essentially demonstrate the information content of the sensors themselves, given the size of assumed model error.

SECTION 5

APPLICATION RESULTS

The theoretical work accomplished in Task 1 of this project and detailed in Appendix A has focussed, primarily, on various approaches to the problem of generating robust residuals and evaluating the associated "information content" for the failure hypothesis testing problems of interest.

In this section, we apply these results to the development of FDI algorithms for sensor bias failures and to the evaluation of Kalman filter based FDI.

5.1 DECISION ALGORITHMS

Although, it is possible to use residuals directly in making decisions, there are practical advantages to using further processing before decisions about a system's failure status is made. For example, residuals can be chosen to represent physical structure alone by ignoring the effects of sensor noise in the metric based optimized parity relation computations. As a result, residuals which are valid over a wider range of operating conditions may be obtained.

We have decomposed the further processing of residuals into two components; information collection and decision logic. The information collection phase is necessitated by the fact that all decisions about the system's operational status must be made by comparing a number (or set of numbers) to a threshold. Thus, all of the failure information must be noninvertibly compressed, over time, into a set of decision statistics upon which the decision

logic will operate to determine the system's status. Furthermore, since the actual performance of the overall FDI algorithm depends highly on its decision statistics, a useful algorithm should allow analytical assessment of these statistics in terms of failure distinguishability. Several candidate algorithms are outlined below.

5.1.1 Maximum Likelihood Decision Rule

Since the information collection algorithm operates on the residuals over time we consider a window of residuals, $\nu_q(k)$ defined by

$$\nu_q(k)^T = [\nu^T(k-q),\ \nu^T(k-q+1),\ \ldots,\ \nu^T(k)] \tag{5-1}$$

where

$$\nu(k) = W_y^T Y_p(k) + W_u^T U_p(k)\ ,$$

and where U_p and Y_p are p-windows of the NC controls and NO outputs respectively and $W = [W_y|W_u]$ is the t by (NO+NC)(p+1) parity check matrix.

Recall now that in our representation of uncertainty, (see Section 4) it turned out that $\nu(k)$, and hence $\nu_q(k)$ could be described statistically by a weighted sum of Guassians (WSOG) probability density function (pdf) for each failure hypothesis. That is;

$$p_{\nu_q}(\zeta|H_i) = \sum_{\ell=1}^{L} P_\ell\, N_{\nu_q}(m_\ell^i\ ;\ P_\ell^i) \tag{5-2}$$

where $N_\zeta(m,P)$ denotes a Gaussian density in the variable ζ with mean, m, and covariance matrix P and p_ℓ denotes the a-priori proability of the ℓth model.

In a classical m-ary hypothesis testing problem, if we can compute $p_{\nu_q}(\zeta|H_i)$, then a simple maximum likelihood (ML) decision rule can be defined. That is,

$$\text{Decide } H_i \text{: if } p_{\nu_q}(\zeta|H_i) > p_{\nu_q}(\zeta|H_j) \text{ , for all } j. \tag{5-3}$$

Although Eq. 5-3 represents an optimal algorithm (in the sense of minimum error probability for the specified pdf's), the computation of these pdf's is quite involved and highly dependent on the set of models which have been assumed. We can, however, simplify the required processing and possibly reduce the sensitivity to the specific set of models by using a Gaussian approximation to the pdf's in Eq. 5-3. One commonly used approximation is the Gaussian density which preserves the first two moments of the original WSOG pdf. The mean and covariance of this approximate Gaussian density are given by,

$$E(\nu_q|H_i) \triangleq m^i = \sum_{\ell} p_\ell \, m_\ell^i \tag{5-4a}$$

$$\text{Var}(\nu_q|H_i) \triangleq C_{\nu_q} = \sum_{\ell} p_\ell \left[P_\ell^i + (m_\ell^i - m^i)(m_\ell^i - m^i)^T\right] \tag{5-4b}$$

Given this approximation, the information collection phase is considerably simplified. Eq. 5-3 reduces to an algorithm where decisions are made on a linear transformation of the windowed residuals [14]. In particular, decisions are based on the log-likelihood-ratio (LLR) statistic,

$$\ell_i = \{(m^i)^T C_{\nu_q}^{-1}\} \nu_q(k) - \frac{1}{2} (m^i)^T C_{\nu_q}^{-1} (m^i) \tag{5-5}$$

In Eq. 5-5, note that m^i and $C_{\nu_q}^{-1}$ are computed off-line and we have assumed that $C_{\nu_q}^{i}$ is the same for all hypotheses as in the case of sensor bias failures. Details of this calculation are now given.

The approximate mean, m^i, is given by,

$$m^i = \sum_{\ell} p_{\ell} \, \overset{i}{m}_{\ell} = E_{\ell} \{\nu_q(k) | H_i\} \tag{5-6}$$

$$\overset{i}{m}_{\ell} = \begin{bmatrix} I_t \\ I_t \\ \vdots \\ I_t \end{bmatrix} \overset{i}{\bar{\nu}}_{\ell} \tag{5-7}$$

where

$$\overset{i}{\bar{\nu}}_{\ell} = [W_y^T \, \bar{Y}_p{}^{i\ell} + W_u^T \, \bar{U}_p{}^{i\ell}] \ ,$$

$$\bar{Y}_p{}^{i\ell} = E \, [Y_p(k) | H_i, \ \ell\text{th model}]$$

$$\bar{U}_p{}^{i\ell} = E \, [U_p(k) | H_i, \ \ell\text{th model}] \quad \text{and}$$

$$I_t = t \times t \text{ identity matrix}$$

The approximate covariance matrix, C_{ν_q} is given by

$$C_{\nu_q} = \sum_{\ell} p_{\ell} \{ C_{\nu_q}^{\ell} + (m_i - \overset{\ell}{m}_i) (m_i - \overset{\ell}{m}_i)^T \} \tag{5-8}$$

$$C_{\nu_q}^{\ell} = \text{Toeplitz} \, [C_\nu{}^0, C_\nu{}^1, \ldots, C_\nu^q] \tag{5-9}$$

$$C_\nu{}^j = W_y{}^T C_{yj} W_y \tag{5-10}$$

$$C_{yj} = \begin{bmatrix} C_i & C_{i+1} & \cdots & C_{i+p} \\ C_{i-1} & C_i & \cdots & C_{i+p-1} \\ \vdots & & & \\ C_{i-p} & C_{i-p+1} & \cdots & C_i \end{bmatrix} \tag{5-11}$$

and where $C_i = C_{-i}{}^T = E\{y(k)\ y^T(k+i)\}$ is computed as described in Appendix A and [13].

Note that, assuming a stationary input, $C_{\nu_q}{}^{\ell}$ is a block Toeplitz matrix (although C_{yj} is not) an hence C_{ν_q} is also Toeplitz. This property is useful since reliable and efficient algorithms exist for computing $C_{\nu_q}{}^{-1}$.

5.1.2 An Alternative FDI Algorithm

The approximate ML algorithm described above represents a considerable simplification over the optimum ML algorithm based on the WSOG density function. In some instances, however, further simplification may be desired. In particular, Eq. 5-5 requires the maintenance of a possibly large window of residuals and the resulting inner product with a large pre-stored sequence. This processing requirement stems, primarily, from the residual sequence being (in general) not "white". If, however, C_{ν_q} were block diagonal ($\nu_q(k)$ uncorrellated or "white"), then the decision statistic, ℓ_i, is based on a simpler computation, viz.,

$$\ell_i \sim \left[\bar{\nu}_i{}^T C_\nu{}^{\circ -1} \right] \sum_{j=0}^{q} \nu(k-j) \tag{5-12}$$

where $\nu_i = E(\nu(k)|H_i)$ and $C_\nu{}^\circ$ is the covariance of $\nu(k)$.

Although, in general, $\nu(k)$ is a correlated sequence, we can define the decision statistic,

$$\lambda_i = g_i^T \sum_{j=0}^{q} \nu(k+j) \tag{5-13}$$

and choose the g_i to maximize the distinguishability of Hypothesis i from other failure hypotheses. That is, we will partition the failure space into sets of pairwise hypotheses and design a single statistic (λ_i) which is useful for distinguishing them. For example, suppose we consider the detection problem in which each failure hypothesis should be as discriminable as possible from normal operation. We can, for example, minimize B_{i0} as discussed in to find the best choice of g_i. For the problem of detecting sensor bias failures,

$$g_i \sim \{ \sum_j \sum_k [C_{\nu_q}]_{jk} \}^{-1} \bar{\nu}_i \tag{5-14}$$

where the term in braces is the covariance matrix of the sum in Eq. 5-13, $[C_{\nu_q}]_{jk}$ is the j-k'th block matrix of dimension t x t in the Toeplitz matrix C_{ν_q} and $\bar{\nu}_i$ is the expected value of $\nu(k)$ under the ith failure hypothesis, (see Eq. 5-7).

In order to define an FDI algorithm which utilizes λ_i, we need to compute the statistics of λ_i. Decision regions (in λ_i - space) are then easily determined using a Gaussian approximation. For the detection problem, we specifically need the mean $E(\lambda_i)$ and variance σ^2_λ, under normal operation (H_0) and under the ith failure (H_i). Specifically,

Under H_0: $$E(\lambda_i) = 0 \tag{5-15}$$

$$\text{var}(\lambda_i) = \bar{\nu}_i^T \{ \sum_j \sum_k [C_{\nu_q}]_{jk} \}^{-1} \bar{\nu}_i$$

(Note here that further simplification of the braced term is possible using the Toeplitz properties of C_{ν_q}).

$$\text{Under } H_i: \quad E(\lambda_i) = g_i^T \cdot \bar{\nu}_i \cdot (q+1) \tag{5-16}$$

$$\text{var}(\lambda_i) = \text{as in 5-15.}$$

A number of observations can be made concerning the Eqs. 5-13 to 5-16.

1. First, any sensor bias failure whose magnitude is greater than that assumed in computing g_i will result in a larger value of $E(\lambda_i)$ under H_i. Thus, the likelihood of making a type 2 error (choosing H_0 under H_i) is smaller for all sensor bias failures which are larger then the design value. This is clearly a desirable property of FDI algorithms.

2. Distinguishability metrics can be computed for λ_i, for each q, thereby indicating the length of the information collection window which is needed to achieve the desired performance. For example, the Bhattacharyya metric, using the approximate Gaussian densities defined by Eqs. 5-15 and 5-16 is,

$$B_{i0} = \frac{1}{8} \bar{\nu}_i^T \left\{ \sum_j \sum_k [C_{\nu_q}]_{jk} \right\}^{-1} \bar{\nu}_i \, (q+1) \tag{5-17}$$

Notice that the dependence on the window size, q, is imbedded in the bracketed term.

5.1.3 Minimizing Decision Delay for Abrubt Failures

The algorithms discussed in 5.1.1 and 5.1.2 are based on moving window calculations for both residual generation and information collection. Implicit in the analysis of these algorithms is the assumption that the effect

of a failure is present throughout the window. In the case of abrupt failures, these algorithms result in a decision delay at least as long as the information collection window. In cases when large uncorrellated errors are present in the residuals (e.g., due to sensor noise) this window can be exceptionally large resulting in possibly large decision delays.

When dealing with abrupt changes in systems a significant problem is the one of unknown onset time. Fig. 5-1 illustrates the improvement in detectability of a constant mean in white Gaussian noise when the optimal processing (maximum likelihood decision statistic) is begun at the failure onset time as opposed to operating over a moving window. The figure suggests that for long windows, one would expect significant performance improvements from the algorithm which is started at the failure onset time.

A variety of techniques have been introduced which attempt to realize performance that approaches the level obtained when failure onset time is known. Willsky's paper [2] provides a good review of the issues and reference [16] is an example of current work in this area.

To avoid the computational burdens which are typical of many solutions to the problem of unknown onset time (e.g., GLR with implicit failure time estimation [2]), we are interested in alternative algorithms. The use of decentralized parity relations of various orders provides us with such an alternative.

First, note that zeroth-order parity relations, that is relations based on memoryless comparison of the values of inputs and outputs at the present time only, respond instantaneously when a failure occurs. Higher-order relations provide information with some time delay. A two-level FDI structure which makes selective use of these relations is now described. First, at the

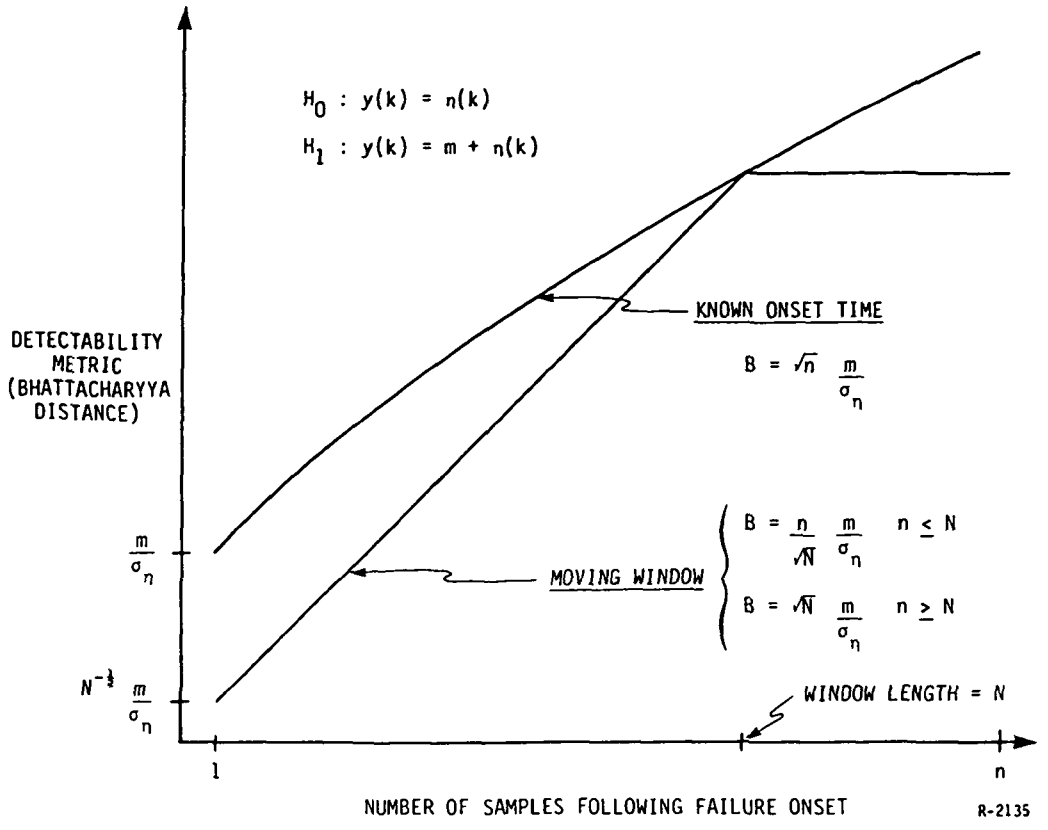


Figure 5-1. Advantage of Known Onset Time

monitoring level we have tests based on the shortest (lowest order) and most sensitive of the parity relations (preferably zeroth order) so that we have full coverage of all failure modes. Alarms are declared at the monitoring level using short intervals for information accumulation and relatively low thresholds. This produces very fast detection at the expense of possibly larger false alarm rates. These alarms, however, are not used as a positive failure indication, but rather as a trigger for the second level. The second level consists of longer running tests (beginning at the initial point in the

monitoring data window if desired, so that no time delay is introduced by the two-step procedure) for reliably identifying the failure mode and rejecting any false start. These somewhat longer tests are based on the same low-order parity relations used at the monitoring level, as well as other higher-order relations.

The second level tests (which are triggered by alarms at the monitoring level) provide final reliable failure decisions. They make use of residuals in a selective way so as to minimize the decision errors and delays of each test. That is, only those residuals which provide reliable information about the hypotheses being tested are used as inputs to these tests. One test which may be employed is known as a Sequential Probability Ratio Test (SPRT). This test has the property that it reaches a decision in as short a time as is possible given the level of uncertainty in the residuals and specified probabilities of correct and incorrect decision.

Fig. 5-2 shows a functional breakdown of an FDI system based on the proposed two level structure for detecting and identifying sensor failures in the F-100 engine example; details of the F-100 model were given in Table 4-1. The trigger mechanism consists of five weighted sum of squared residuals (WSSR) statistics (one for each failure mode) which are used in parallel to generate a single trigger. Each WSSR statistic, s_k, is defined by

$$s_{k+1} = \alpha\, s_k + (1-\alpha)\, W_k \qquad (5\text{-}18)$$

where,

$$W_k = \nu(k)^T C_\nu^{-1} \nu(k),$$

Only those residuals which contribute significantly to the distinguishability of a single failure from normal operation and respond quickly to failures (i.e., low order parity checks) are used in each trigger test. The parameter α

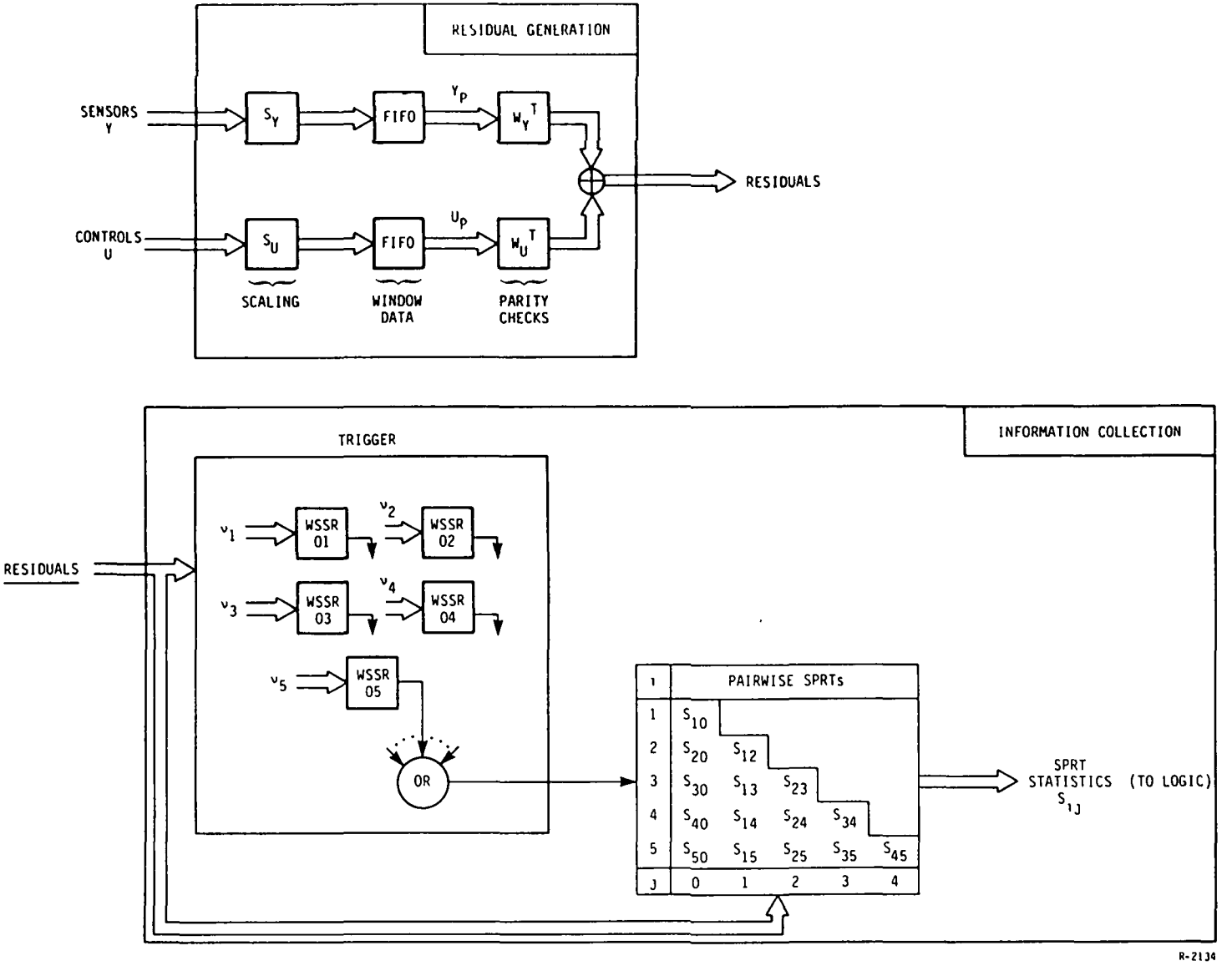


Figure 5-2. Functional Breakdown of FDI Algorithm

is chosen on the basis of tradeoffs between decision delay and probability of false triggering. The individual thresholds for each test are chosen based on Chi-squared distributions and the desired probability of missing a failure in the trigger mechanism.

Following a trigger, 15 SPRT tests are initiated (one for each pair of failure hypotheses; five sensor failures plus the no-fail hypothesis). For sensor bias failures, we use an SPRT statistic, s_k, which is defined by the log-likelihood-ratio for two means in white noise and is given by,

$$s_k = s_{k-1} + [\bar{\nu}_i - \bar{\nu}_j]^T C_\nu^{-1} \quad \nu(k) - \frac{1}{2} [\bar{\nu}_i^T C_\nu^{-1} \bar{\nu}_i - \bar{\nu}_j^T C_\nu^{-1} \bar{\nu}_j] \quad (5\text{-}19)$$

Only those residuals which contribute significantly to the distinguishability of failure mode i from failure mode j, (note: $\bar{\nu}_0 = 0$) are used in each SPRT. The SPRT algorithm chooses hypothesis i over hypothesis j if $s_k > t^+$, hypothesis j over i if $s_k < t^-$, and continues sampling when $t^- < s_k < t^+$. A decision is made when all the SPRT's "vote" in favor of a single hypothesis over all others. When a decision event doesn't occur, we are left with an indication of the ambiguity set which the algorithm can not resolve using the current data. Finally, note that in Eq. 5-19, the size and sign of the sensor bias failure is assumed known. When neither size or sign is known, the SPRT's must be modified so that any triggerable failure is correctly isolated; e.g., see [20].

5.1.4 Simulation Results

The algorithm of Fig. 5-2 was coded and tested on a linear simulation of the F-100 engine at h = 10k, M = .6, PLA = 83°. Five zeroth order robust detection parity checks using the "null space" approach (Section 4) were generated based on model uncertainties corresponding to the curve fit errors [11]

of the reduced order linear model being used and subsets of these selected for each test. The "jth" WSSR and "jth" SPRT tests for distinguishing failures from normal behavior used only a single parity check (the one corresponding to sensitivity to the jth failure) while the "i-jth" SPRT tests for distinguishing between failure i and failure j used two parity checks (the ones corresponding to sensor failure i and sensor failure j sensitivity). Table 5-1 shows the values of the coefficients which multiply scaled versions of the measurements and control values for each parity check. The number in the metric column is related to the robustness of the parity check, where zero is a perfect parity check and each metric must be smaller than 5.6 (largest eigenvalue of the "C_0" matrix in Section 4). Notice that failures of sensor 1 and 5 show up in parity checks one and five in a very similar manner. This implies a possible difficulty in distinguishing these two failures.

Figure 5-3 shows the five residuals (scaled by 1000) obtained from the corresponding parity checks and noiseless measurements generated by the linear F-100 engine model with fuel flow as the only input (modeled as a first order Markov process with 0.1 sec. time constant and standard deviation of 350 PPH). Notice that even without sensor noise, the residual is not identically zero. Variations in the state modulate the modeling errors through the imperfect parity checks. A 50 rpm bias is abruptly introduced at t = 1.0 sec. and shows up, as expected, primarily in residual numbers one and five, (see Table 5-1). Figure 5-4 shows the five WSSR statistics with $\alpha = .37$ for each test.

Clearly both WSSR statistics one and five can trigger SPRT tests when a failure of sensor 1 occurs. Fig. 5-5 shows the SPRT tests designed to distinguish each failure from no failure (s_{i0}, i=1,..., 5). The WSSR thresholds were chosen so that a false trigger was initiated at .42 seconds.

TABLE 5-1
ZEROth ORDER ROBUST DETECTION PARITY CHECKS
(Null Space Approach)

Sensor Sensitivity Constraint	p	Metric	Scaled Sensor Coefficients					Scaled Control Coefficients				
			N1	N2	PT4	PT6	FTIT	WF	AJ	FGV	SVA	BLC
1	0	4.487E-02	0.6828	0.2472	0.0656	-0.0023	0.4456	-0.4577	-0.2341	0.0080	0.0013	-0.0678
2	0	2.231E-02	0.1357	0.6185	-0.5738	0.0063	0.1052	0.0247	-0.2999	0.0168	0.0028	-0.4097
3	0	2.266E-02	0.0350	-0.5573	0.6469	0.0023	0.0194	-0.1659	0.2625	-0.0164	-0.0027	0.4154
4	0	8.543E-01	-0.0508	0.2525	0.0944	0.5901	0.0682	-0.1552	0.3104	-0.0268	-0.0002	-0.6714
5	0	4.391E-02	0.6411	0.2752	0.0523	0.0044	0.4653	-0.4741	-0.2471	0.0085	0.0014	-0.0895

Scale Factors Definitions

$$\nu = W_y^T y + W_u^T u$$

$$y = S_y \cdot Y$$

$$u = S_u \cdot U$$

$$S_y = \text{diag}\,[1E4,\ 1.5E4,\ 550,\ 130,\ 1600]$$

$$S_u = \text{diag}\,[1.5E4,\ 5,\ 50,\ 10,\ 1]$$

Following the false trigger, all SPRTS in Fig. 5-5 indicate that no failure has occurred by t = 0.70 seconds by crossing the threshold $t^- = -9.2$, and are therefore turned off. The thresholds ($t^+ = 9.2$ and $t^- = -9.2$) correspond to equal probabilities of type 1 and type 2 decision error and are determined from equations found in [2] and [6]. Finally, Fig. 5-6 shows the five SPRT tests for distinguishing sensor number 1 failures from all other hypotheses.

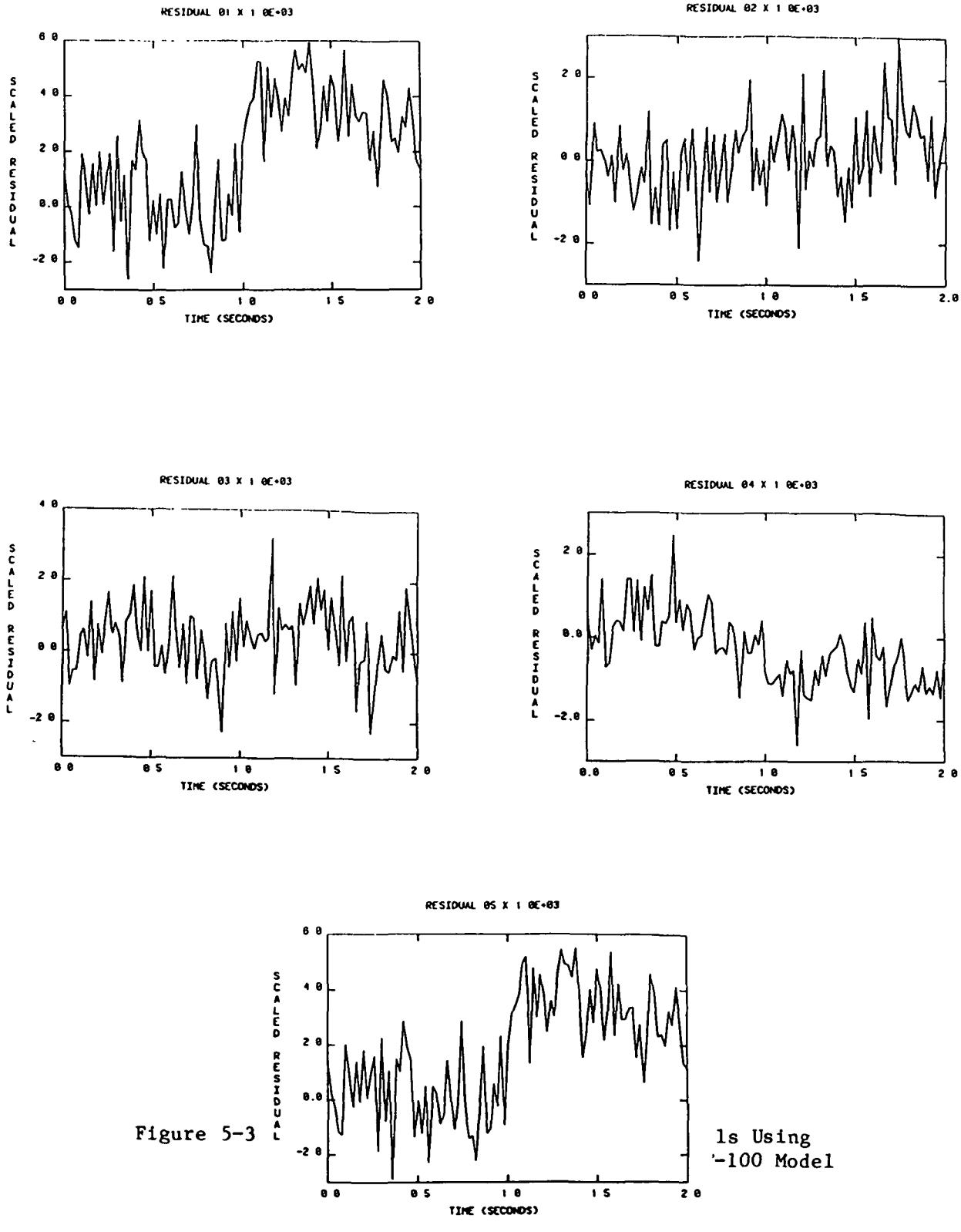


Figure 5-3 ls Using
-100 Model

Figure 5-3. Five Zeroth Order Detection Residuals Using Noiseless Measurements and Linear F-100 Model

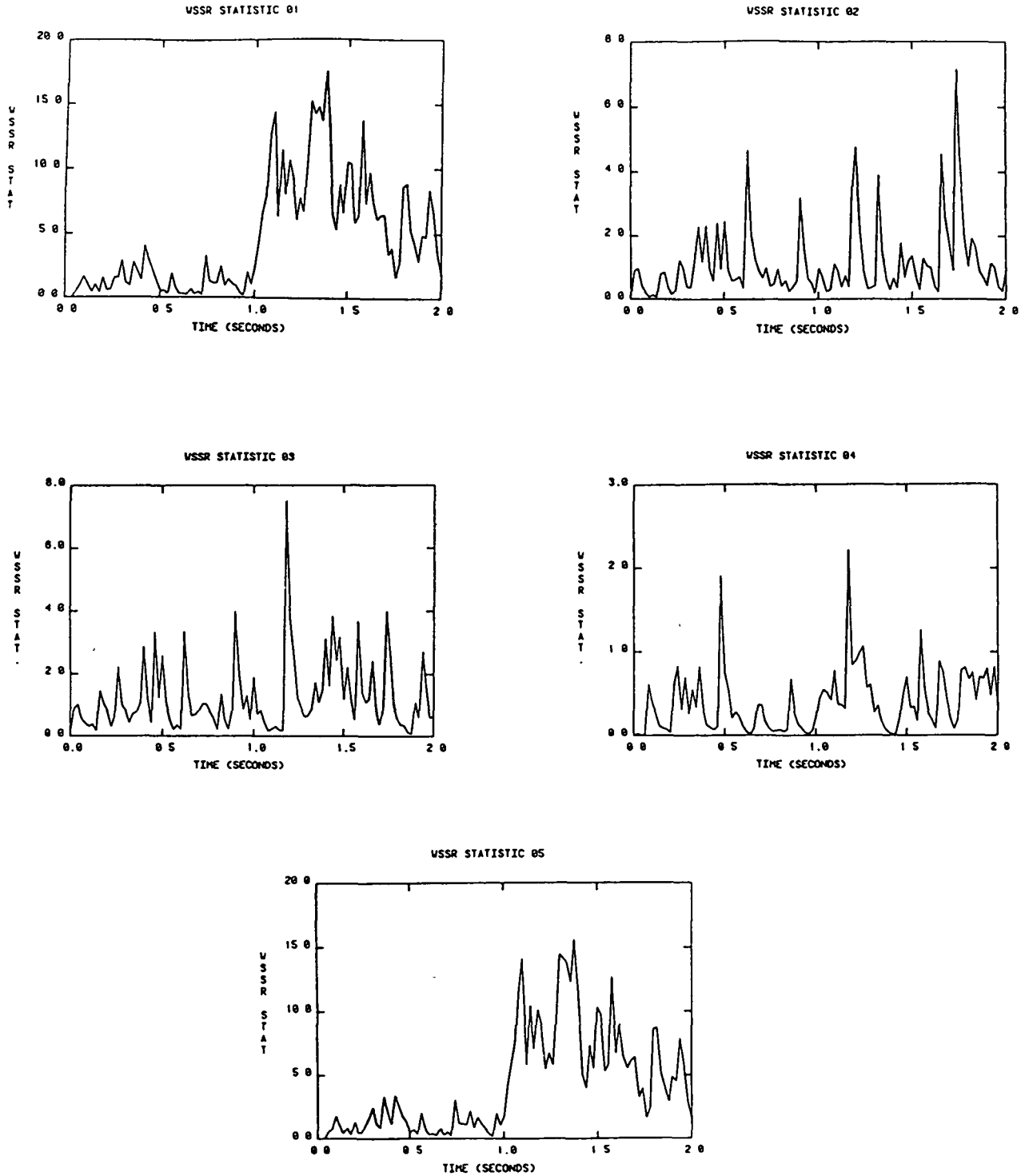


Figure 5-4. WSSR Statistics Corresponding to Residuals in Figure B-3.

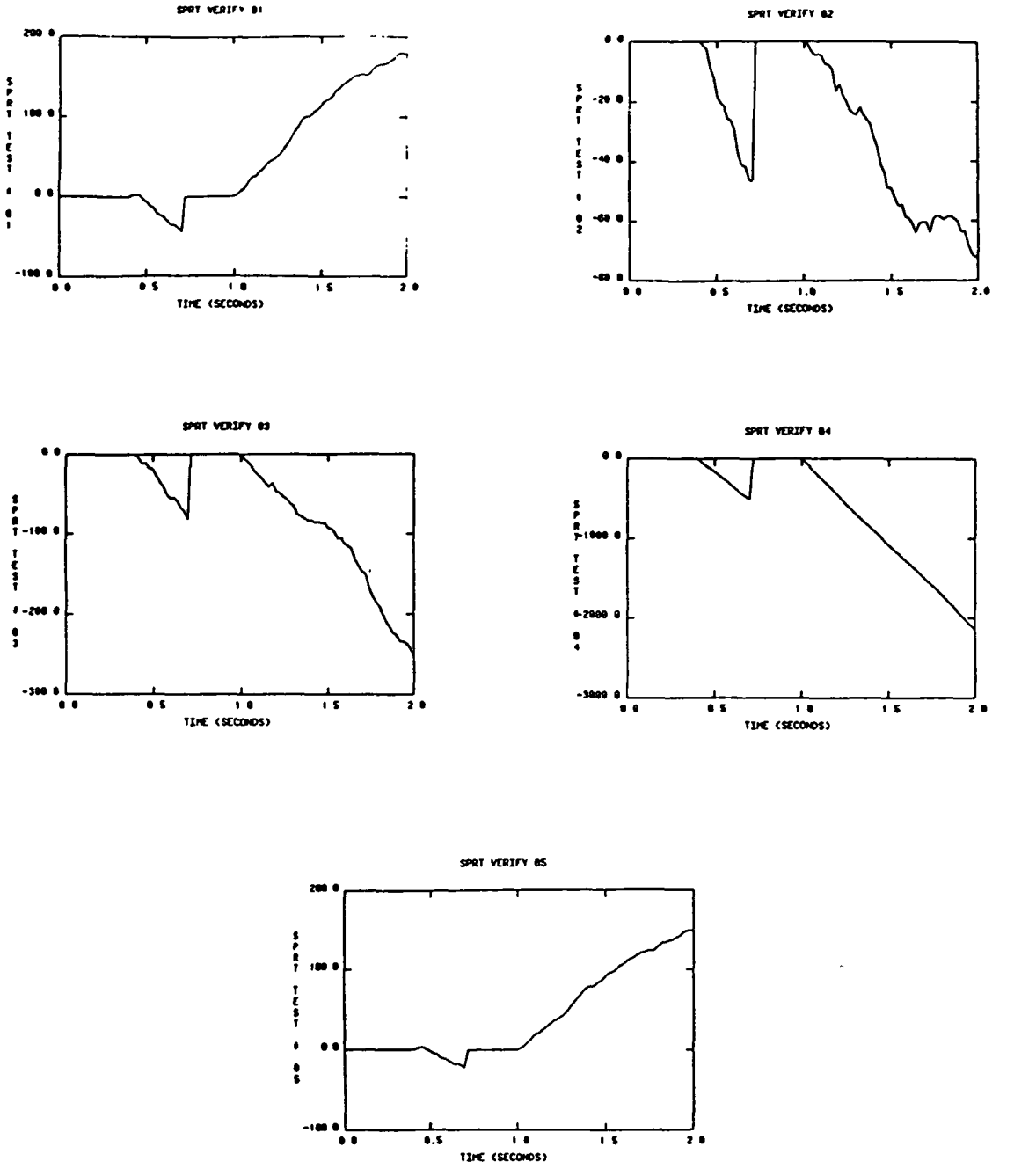


Figure 5-5. SPRT's for Distinguishing Sensor-i Failures from Normal Operation

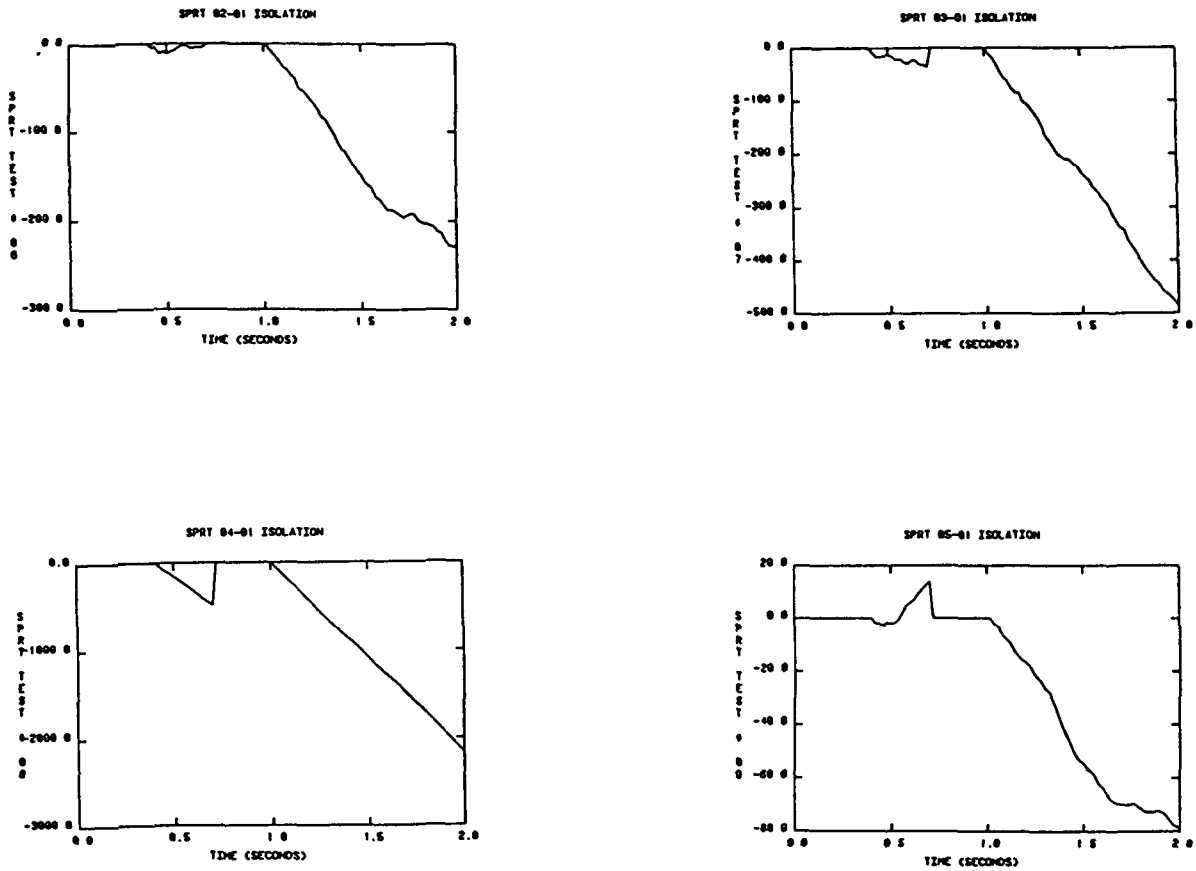


Figure 5-6. SPRT's for Distinguishing Sensor 1 Failures

As expected the most difficult decision (and hence longest SPRT decision delay) is between sensor #1 and #5. However a correct isolation of sensor failure #1 is obtained at t = 1.24 seconds when $s_{10} > t^+$, and s_{21}, s_{31}, s_{41}, $s_{51} < t^-$.

In summary, we have presented three potential decision algorithms for detecting and isolating sensor bias failures in the F-100 jet engines at a single operating point. One of these was chosen and implemented in Fortran code (see Appendix C) and demonstrated with a linear simulation of the F-100 engine.

In the next subsection, we apply the theoretical results of Section 4 to the evaluation of Kalman filter based algorithms.

5.2 KALMAN FILTER EVALUATION

As discussed in Section 4 and Appendix A, the metrics we have proposed for use in selecting parity checks can also serve as an evaluation tool for any FDI system. Task 3 of this project consists of evaluating FDI algorithms based on Kalman filter residuals and in particular, the algorithm discussed in reference [3].

The algorithm reported in [3] consists of a single Kalman filter (KF) which uses all five available measurements to estimate the four state variables (see Table 4-1 for model details) and produce a vector of five residuals. These residuals are then used to form a weighted sum of squared residuals (WSSR) detection statistic. The WSSR is summed over a finite window, and compared to a single threshold for detecting the presence of a sensor failure. Isolation of the failed sensor is accomplished through the use of five KF's ($KF_1,\ldots,KF_5$), each of which makes use of only four of the

available measurements. The residuals from the five isolation filters are then used to form 5 WSSR statistics which are then low-pass-filtered (a recursive, exponential age weighting of the WSSR statistic) and compared to the statistic generated by a KF which uses all five measurements, (KF_0). Ideally, when sensor i fails, the statistics produced by KF_0 and KF_j, $j \neq i$, will "diverge" due to the presence of a failed sensor in each of these filters. The statistic produced by KF_i will remain comparatively small.

A number of error sources are possible contributors to poor performance of this algorithm. First, in the detection phase, modeling error produces non-ideal residuals, (namely nonwhite), thus causing a decrease in detection performance from that which could have been predicted on the basis of the chosen WSSR weights and thresholds. The exact nature of how modeling error can effect performance depends on the choice of Kalman Gain as well. Thus, our metric based approach to assessing the impact of modeling error on KF-based FDI algorithms depends highly on the choice of Kalman Gain.

To demonstrate the use of metrics in evaluating Kalman Filter based algorithms, we designed a Kalman Gain using "LQGALPHA," a software package developed by ALPHATECH which includes the capability of computing steady state Kalman Gains from specified process and sensor noise covariances matrices.

The standard deviations for the corresponding sensor noises were as follows:

$$\sigma_{N1} = 13.9 \text{ RPM}$$

$$\sigma_{N2} = 12.1 \text{ RPM}$$

$$\sigma_{PTR} = 0.67 \text{ PSI}$$

$$\sigma_{PT6} = .09 \text{ PSI}$$

$$\sigma_{FTIT} = 2.2^\circ \text{ F}$$

A fuel flow variance of 350 PPH, (all other control inputs assumed zero) was used. These noise variances were then scaled in accordance with the scaling (both output scaling and "correction factors") applied to the nominal model. The resulting gain (discrete time KF formulation) is shown in Table 5-2. The equivalent continuous time filter has a transfer function (states to state estimates) with eigenvalues of about -40, -6, -2, -.8 (rad/s.). Note that the mode corresponding to the smallest eigenvalue (-40) has a time constant on the order of the sampling time (T = .02 sec) indicating that certain linear combinations of the measurements are used directly at each time step with little or no filtering.

TABLE 5-2. KALMAN GAIN AND EQUIVALENT CONTINUOUS TIME TRANSFER FUNCTION (STATES TO ESTIMATE) AT "DC"

Discrete Time Steady State Kalman Gain				
1.4870000E-01	2.1830000E-01	7.4580000E-02	1.4280000E-01	-2.7160000E-01
7.2420000E-02	1.0640000E-01	3.6350000E-02	6.9550000E-02	-1.3230000E-01
4.8110000E-03	7.0520000E-03	2.4070000E-03	4.6140000E-03	-8.7930000E-03
5.8830000E-03	8.6410000E-03	2.9500000E-03	5.6430000E-03	-1.0750000E-02

Figure 5-7 shows the five KF residuals of the nominal KF (Table 5-2) with an N1-bias failure of 50 RPM occurring at t = 1.0 seconds. The truth model used to generate the residuals in Fig. 5-7 was chosen from the set of randomly perturbed system matrices used in the robust parity check design process. Software for implementing this process is described in Appendix E. The system matrices were generated by adding zero-mean Gaussian errors to the nominal system matrices of Table 4-1. The variances of the individual terms were

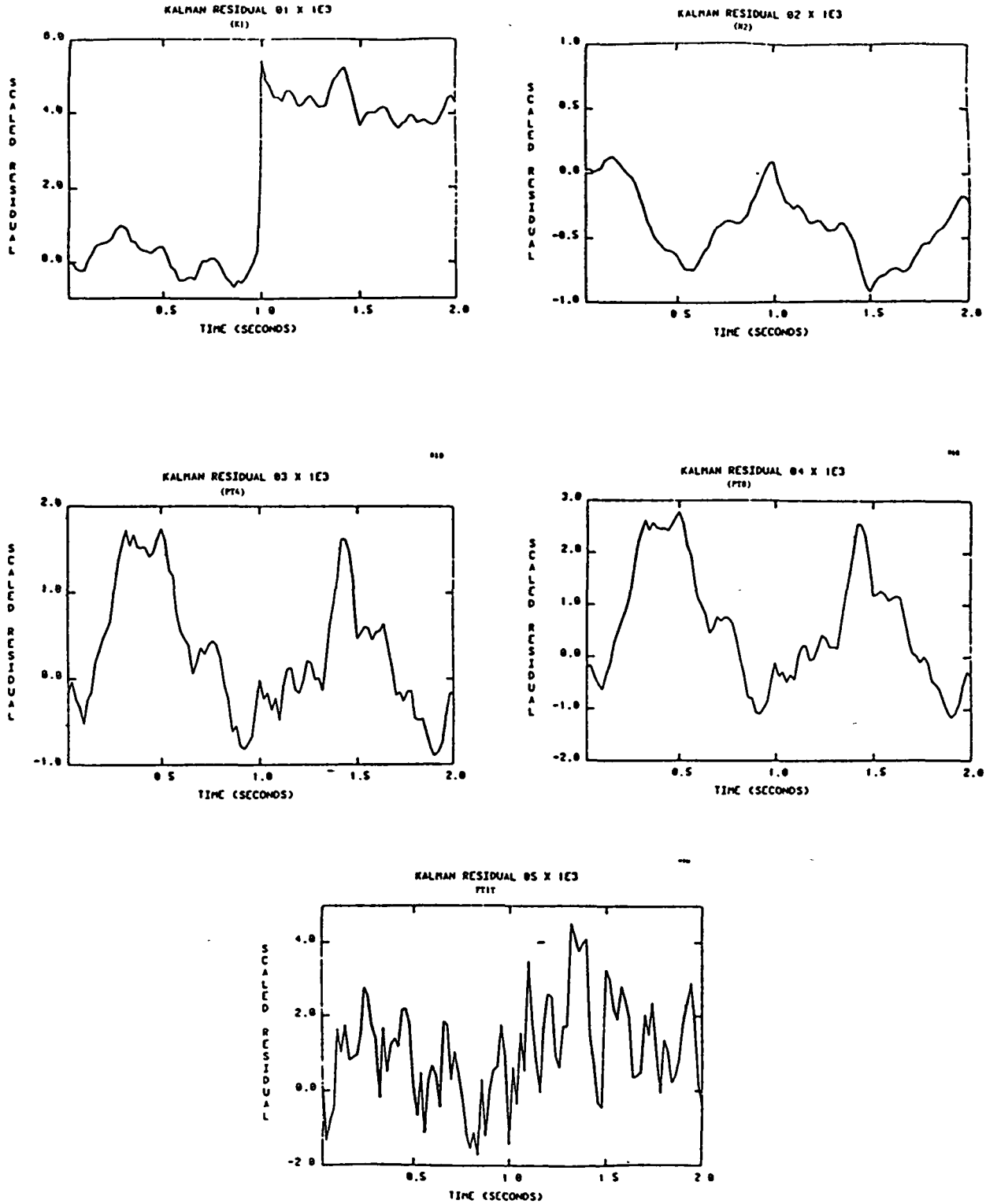


Figure 5-7. Kalman Filter Residuals with 50 rpm N1-bias Failure at t = 1.0 sec.

obtained from the reduced order model curve fit errors ("$\sigma_{y/x}$") reported in [11] (and shown in Table 4-1), and checks for outliers are performed before using any result. No sensor noise is simulated since the realistic values of noise variances tend to obscure the robustness results. A 1st order Markov model for fuel flow is used as the input with a correlation time of 0.1 seconds and a steady state standard deviation corresponding to 350 PPH. This input is intended to model the correlation structure of the input generated by the control system which regulates the engine at this operating point.

With no sensor noise and measurable control inputs, an ideal filter would have residuals which are identically zero during no failure and a bias in each residual following the failure. Thus the effects of model mismatch are easily derived from Fig. 5-7. The expected residual bias (scaled by 1000 as in the figure) due to an N1-failure can be calculated using the KF evaluation results discussed subsequently and in Appendix A. In particular, $E\{\nu_{KF}\} = [4.1, -0.4, -0.3, -0.2, 1.5]^T$. As is evident in the figure, deterioration of the KF residuals due to modeling error is not particularly severe in terms of detecting an N1-sensor failure if one allows for the non-ideal behavior in setting thresholds.

5.2.1 Evaluation of KF Using the Bhattacharyya Distance

Characterization of KF performance cannot, of course, be realistically inferred from a single simulation. The use of performance metrics, however, can provide a realistic characterization by assessing the ensemble-averaged impact of model errors in terms of FDI performance. First, however, we must extend the results of Appendix A to include systems with known control inputs. Consider a Kalman Filter which is based on the nominal system,

$$x(k+1) = A_o\, x(k) + B_o\, u(k) \tag{5-20}$$

$$y(k) = C_o\, x(k) + D_o\, u(k) \tag{5-21}$$

The mean and covariance of the residuals for the nominal KF under the ith failure and the ℓth model hypothesis $(A_\ell, B_\ell, C_\ell, D_\ell)$ are computed, from the dynamic equations,

$$x_a(k+1) = F_\ell\, x_a(k) + G_\ell\, u(k) + \bar{K}\,(v_k + b_i) \tag{5-22}$$

$$\nu_{i_\ell}(k) = H_\ell\, x_a(k) + M_\ell\, u(k) + v_k + b_i \tag{5-23}$$

where, we have,

$$F_\ell = \begin{bmatrix} A_\ell & 0 \\ A_o K_o C_\ell & A_o[I - K_o C_o] \end{bmatrix} \tag{5-24}$$

$$G_\ell = \begin{bmatrix} B_\ell \\ B_o + A_o K_o [D_\ell - D_o] \end{bmatrix} \tag{5-25}$$

$$\bar{K} = \begin{bmatrix} 0 \\ A_o \quad K_o \end{bmatrix} \tag{5-26}$$

$$H_\ell = [\, C_\ell \mid -C_o \,] \tag{5-27}$$

$$M_\ell = D_\ell - D_o \tag{5-28}$$

Thus, under the ith failure and ℓth model we have,

$$\bar{\nu}^{i\ell} \triangleq E(\nu|H_i,\ell) = H_\ell \bar{x}_a^{i\ell} + M_\ell \bar{u} + b_i \tag{5-29}$$

where

$$\bar{x}^{i\ell} = [\ I - F_\ell\ ]^{-1}\ \{\ G_\ell \bar{u} + \bar{K}\ b_i\ \} \tag{5-30}$$

and

$$C_a^\ell \triangleq \text{cov}(\nu|H_i,\ell) = F_\ell\ C_a^\ell\ F_\ell^T + G_\ell\ Q\ G_\ell^T + \bar{K}\ R\ \bar{K}^T \tag{5-31}$$

where

$$R = \text{cov}(v_k)$$

$$Q = \text{cov}(u_k)$$

Note that, in Eqs. 5-20 and 5-21, the measurable input sequence is assumed to be a white noise process. Other statistical characterizations of the input sequence can be included by adding additional states to Eq. 5-22. Highly correlated input sequences tend to be the "worst case" for evaluation purposes since the modeling error (which is modulated by the state) would then tend to produce highly correlated residuals and be most confused with sensor biases.

These results were applied to the KF described by the Kalman gain in Table 5-2 in order to assess the mismatched-model-ensemble-average FDI performance. Table 5-3 shows the pairwise B-distances between sensor bias failure modes (positive and negative failures of sensors; i = 1,...,5, and no failure; i = 0), based on a Gaussian approximation to the "weighted sum of (10) Gaussians" probability density function which is implied by our multiple truth model assumptions and the input and sensor noise covariances used to generate the gain in Table 5-2. In the table, larger distances imply lower achievable misclassification rates when the assumed Gaussian statistics are

TABLE 5-3. BHATTACHARYYA DISTANCES FOR SENSOR BIAS FAILURES KALMAN FILTER RESIDUALS

SENSOR	FAILURE	0	1	2	3	4
N1	1 +	1.111E+00				
	-	1.111E+00				
N2	2 +	1.562E+00	3.321E+00			
	-	1.562E+00	2.024E+00			
PT4	3 +	2.146E+00	3.581E+00	4.238E+00		
	-	2.146E+00	2.933E+00	3.179E+00		
PT6	4 +	9.798E+00	1.109E+01	1.165E+01	1.165E+01	
	-	9.798E+00	1.073E+01	1.107E+01	1.224E+01	
FTIT	5 +	4.600E-01	6.593E-01	1.258E+00	2.170E+00	9.861E+00
	-	4.600E-01	2.482E+00	2.786E+00	3.042E+00	1.065E+01

used. Note that the B-distance for detecting FTIT sensor bias failures is quite small indicating that the assumed failure size of 8°F may not be easily detectable. Other failure sizes assumed in Table 5-3 correspond to 50 RPM for N1 and N2 sensors and 3 PSI for PT4 and PT6. In addition, the B-distance between positive bias failures of N1 and FTIT is also small enough to indicate potential misclassification problems.

The separate effects of sensor noise and modeling error can easily be assessed by setting the sensor noise equal to zero and computing the pairwise B-distances. These distances are given in Table 5-4. All distances are larger than in Table 5-3 as expected although the relative ordering changes somewhat.

Finally, Table 5-5 shows distances for a perfectly known model with nonzero sensor noise. Comparing Table 5-5 with Tables 5-3 and 5-4 we can conconclude that deterioration of the filter (smaller distances) due to modeling error is not particularly severe in terms of potential FDI performance.

TABLE 5-4. BHATTACHARYYA DISTANCES FOR SENSOR BIAS FAILURES
KF WITH NO SENSOR NOISE

SENSOR FAILURE	0	1	2	3	4
1 +	1.680E+01				
-	1.680E+01				
2 +	2.197E+01	5.801E+01			
-	2.197E+01	1.953E+01			
3 +	2.591E+01	4.274E+01	5.754E+01		
-	2.591E+01	4.269E+01	3.824E+01		
4 +	1.396E+01	2.545E+01	4.264E+01	3.927E+01	
-	1.396E+01	3.608E+01	2.923E+01	4.049E+01	
5 +	4.275E+00	1.031E+01	2.200E+01	2.401E+01	1.708E+01
-	4.275E+00	3.185E+01	3.050E+01	3.636E+01	1.940E+01

TABLE 5-5. BHATTACHARYYA DISTANCES FOR SENSOR BIAS FAILURES
KF WITH NO MODELING ERRORS

SENSOR FAILURE	0	1	2	3	4
1 +	1.294E+00				
-	1.294E+00				
2 +	1.864E+00	3.731+00			
-	1.864E+00	2.583E+00			
3 +	2.395E+00	4.006E+00	4.564E+00		
-	2.395E+00	3.372E+00	3.955E+00		
4 +	1.313E+02	1.352E+02	1.357E+02	1.351E+02	
-	1.313E+02	1.300E+02	1.307E+02	1.324E+02	
5 +	6.828E-01	8.611E-01	1.480E+00	2.490E+00	1.272E+02
-	6.828E-01	3.092E+00	3.614E+00	3.667E+00	1.368E+02

However, it should be noted that the major effect of modeling errors (Table 5-5 vs. 5-3) appears to be in the distances involving sensor number four (PT6), suggesting possible filter error and FDI problems for filters which rely heavily on PT6 for updating state estimates.

5.2.2 Extension of the Evaluation Technique for Unknown Base Points

The results above suggest that error sources other than parametric modeling error may be the primary cause of poor performance of the algorithm in [3]. One source which has been neglected until now is the uncertainty in the linearization or base-point of the linearized system model. To include this error source in our analysis, we proceed as follows.

$$\delta x(k+1) = A_o\ \delta x(k) + B_o\ \delta u(k) \tag{5-32}$$

$$\delta y(k) = C_o\ \delta x(k) + D_o\ \delta u(k) \tag{5-33}$$

where δx, δy, and δu represent deviations from a nominal set of basepoints, X_B°, Y_B° and U_B°. The statistics of the KF residuals under the ℓth model hypothesis (A_ℓ, B_ℓ, C_ℓ, D_ℓ, $X_B{}^\ell$, $Y_B{}^\ell$, $U_B{}^\ell$) and the ith bias failure hypothesis ($y = C_\ell x + D_\ell u + b_i$) can then be computed from the dynamic equations;

$$x_a(k+1) = F_\ell\ x_a(k) + G_\ell\ \delta u_k + \bar{K}(v_k+b_i) + b_a \tag{5-34}$$

$$\nu(k) = H_\ell\ x_a(k) + M_\ell\ \delta u(k) + v_k + b_i + b_\nu \tag{5-35}$$

where F_ℓ, G_ℓ, H_ℓ, D_ℓ, and $\bar{K}$ are the same as in Eq. C-5 through C-9, and;

$$b_a = \begin{bmatrix} [I-A_\ell]\ [X_B{}^\ell - X_B^\circ] - B_\ell\ [U_B{}^\ell - U_B^\circ] \\ A_o K_o[(Y_B{}^\ell - Y_B^o) - C_\ell\ (X_B{}^\ell - X_B^o) - D_\ell\ (U_B{}^\ell - U_B^o)] \end{bmatrix}$$

$$b_\nu = (Y_B{}^\ell - Y_B^o) - C_\ell\ (X_B{}^\ell - X_B^o) - D_\ell\ (U_B{}^\ell - U_B^o)$$

The nominal base points for the F-100 engine model (Table 4-1) correspond to the unscaled values shown in Table 5-6. Clearly, comparing the bias sizes which we are evaluating (see Table 5-7) with Y_B^o, even a 1% error in base point determination will result in severe degradation of FDI performance. Table 5-7 shows the degradation in terms of the Bhattacharyya distance metric for a 1% error perturbation of all base points. Comparing Tables 5-7 and 5-3, we see that base-point uncertainty can result in up to a 3-order-of-magnitude decrease in pairwise B-distances. Not surprisingly, the B-distances associated with sensor 4 (PT6) are affected least since the nominal base point (and hence the 1% perturbations) for this sensor is smallest (in comparison to its corresponding failure magnitude). Thus a preliminary conclusion that can be drawn regarding the use of this KF for FDI is that base point accuracy plays a major role in defining the limits of FDI performance.

TABLE 5-6. BASE POINT VALUES AND SENSOR FAILURES

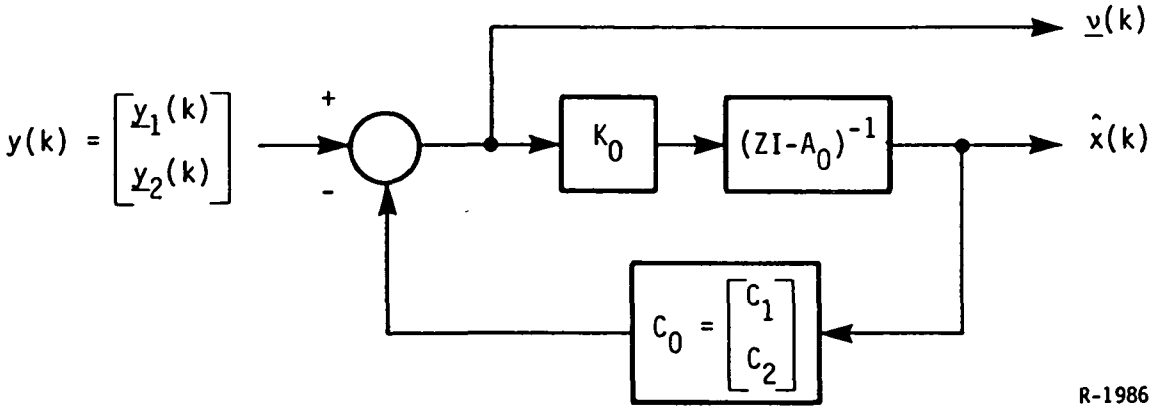
X_B^o =	[8900 rpm, 11,700 rpm, 186° F, 139°F]
U_B^o =	[8000 PPH, 3 ft.^{2}, -25 deg., 3 deg., 0%]
Y_B^o =	[8900 rpm, 11,700 rpm, 284 PSI, 42.5 PSI, 1343° F]
SENSOR Bias Failure Magnitudes	[50 rpm, 50 rpm, 3 PSI, 3 PSI, 8°F]
1% errors in Y_B^o	[89rpm, 117rpm, 28.4PSI, 4.25PSI, 13.4°F]

TABLE 5-7. BHATTACHARYYA DISTANCES FOR SENSOR BIAS FAILURES
KF WITH 1 % UNCERTAIN BASE POINTS

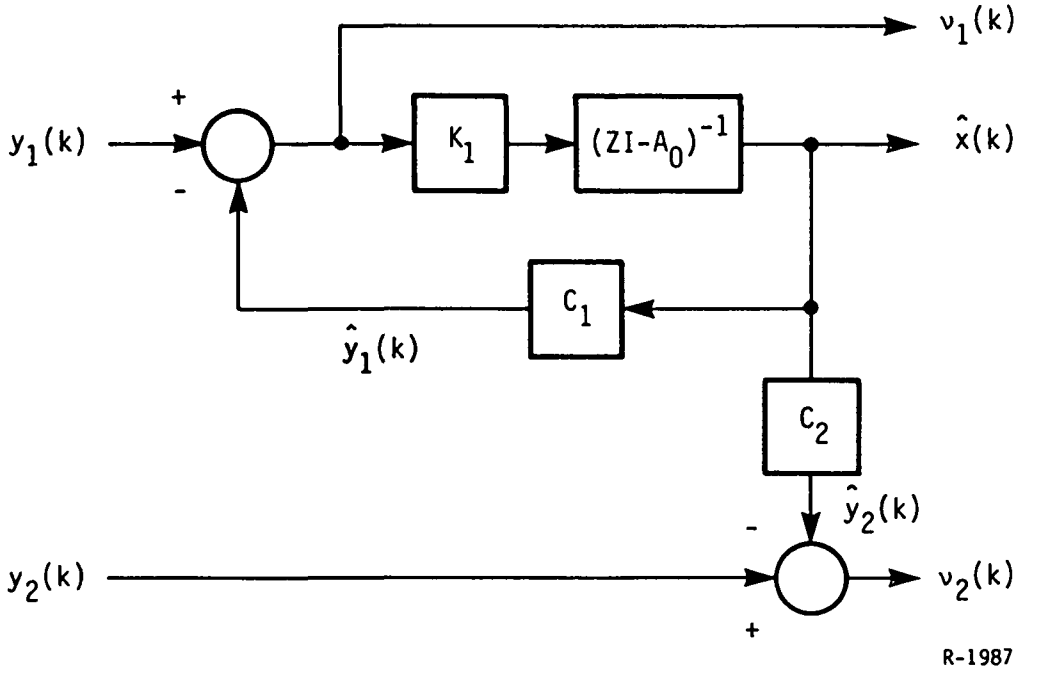
SENSOR FAILURE		0	1	2	3	4
1	+	2.149E-02				
	-	6.902E-02				
2	+	6.291E-02	6.557E-02			
	-	1.659E-01	2.193E-01			
3	+	1.358E-01	1.781E-01	2.114E-01		
	-	9.107E-02	1.050E-01	2.099E-01		
4	+	2.799E+00	2.740E+00	2.867E+00	3.652E+00	
	-	2.812E+00	2.928E+00	2.939E+00	2.152E+00	
5	+	3.063E-02	2.043E-02	3.675E-02	1.548E-01	2.778E+00
	-	8.625E-02	1.526E-01	2.745E-01	1.546E-01	2.915E+00

5.2.3 Results for KF Improvement

FDI schemes which rely on residuals from a Kalman Filter are fundamentally "centralized" approaches. That is, a large dynamic system model is developed and all information relating to that model is used without regard to the quality of the corresponding model. Efforts to tune a Kalman filter based on known uncertainties can be effective, but these efforts are largely heuristic in nature and no well-defined and quantifiable methodology exists. The generation of residuals through robust parity checks is fundamentally a "decentralized" approach to FDI. That is, only the best relationships between measured variables are selected and used individually in the FDI process. Thus, one might expect that improvements to the Kalman filter based approach might be achieved if some sort of partial decentralization could be performed. Figure 5-8 gives one example of a method for partial decentralization. In Fig. 5-8a, the measurements, y(k), are decomposed into two sub-vectors $y_1(k)$ and $y_2(k)$. When the model for y_2 is not well known for example, severe



5-8a. Nominal Filter



5-8b. Partially Decentralized Filter

Figure 5-8. Partial Decentralization of a Kalman Filter

estimation errors may occur in the nominal filter as the "good" information (y_1) is mixed with the "bad", (y_2) through the nominal gain matrix K_o. In Fig. 5-8b filtered state estimate are formed using only $y_1(k)$. The resulting estimates, (which are not corrupted by the inaccurately modeled measurements $y_2(k)$), could then be used to generate residuals corresponding to the measurements $y_2(k)$. Of course, if one uses the ill-modeled description of $y_2(k)$ (namely C_2) the residuals, ν_2, will contain errors. Alternatively, one might consider using $y_2(k)$ and $y_1(k)$ in a robust parity relation to form additional residuals or estimates of y_2.

In order to determine the potential for partial decentralization for the F-100 engine model we considered the use of metrics for FDI in determining the quality of various parts of the overall model. In particular, we considered five multi-input-single-output (MISO) systems corresponding to each of the five measurable quantities (N1, N2, PT4, PT6, FTIT). We then generhigh order (p = 5) robust parity checks for each of these systems using the robust-redundancy null-space approach (see Appendix E). This approach, effectively, gives the best Auto-Regressive-Moving-Average (ARMA) model for the single output system from the set of randomly perturbed models we have defined In addition, the approach provides a metric for comparing the quality of these ARMA models (the smallest eigenvalue of the symmetrized partial observability matrix). The results are shown in Table 5-8. Comparisons of the metrics corresponding to the best ARMA model (parity check) for each MISO system suggests that the models for PT4 and FTIT are particularly poor in comparison to the other sensors. Thus, it is possible that improved filter performance may be achieved by removing PT4 and FTIT sensors from the filter. Of course, estimates of PT4 and FTIT may be needed for the overall control law, in which

case the robust ARMA models of Table 5-8 or those generated through other optimization strategies (see Appendix B) may be used.

TABLE 5-8. *EIGENVALUES AND ARMA COEFFICIENTS* (p = 5) FOR MISO SYSTEMS

SENSOR	METRIC	SCALED AR COEFFICIENTS	SCALED MA COEFFICIENTS				
	3.710E-05	-0.1137	-0.0057	-0.0081	0.0020	-0.0004	0.0045
		0.3850	0.0130	0.0184	-0.0046	0.0009	-0.0103
N1		-0.5628	-0.0139	-0.0196	0.0049	-0.0010	0.0110
		0.5669	0.0132	0.0185	-0.0046	0.0009	-0.0104
		-0.4210	-0.0067	-0.0093	0.0023	-0.0004	0.0052
		0.1457	0.0000	0.0000	0.0000	0.0000	0.0000
	4.770E-06	-0.1192	-0.0028	-0.0019	0.0001	0.0003	0.0015
		0.3921	0.0062	0.0042	-0.0003	-0.0006	-0.0035
N2		-0.5645	-0.0066	-0.0045	0.0003	0.0007	0.0037
		0.5671	0.0063	0.0043	-0.0003	-0.0006	-0.0035
		-0.4148	-0.0031	-0.0021	0.0002	0.0003	0.0017
		0.1393	0.0000	0.0000	0.0000	0.0000	0.0000
	1.050E-02	0.0896	-0.0190	0.0394	-0.0026	-0.0005	0.0671
		-0.2932	0.0640	-0.1278	0.0084	0.0015	-0.2202
PT4		0.4196	-0.0929	0.1821	-0.0120	-0.0019	0.3158
		-0.4213	0.0935	-0.1827	0.0120	0.0019	-0.3172
		0.3093	-0.0695	0.1336	-0.0088	-0.0013	0.2333
		-0.1041	0.0240	-0.0446	0.0030	0.0004	-0.0788
	8.881E-04	0.0734	-0.0082	0.0403	-0.0034	0.0001	-0.0889
		-0.2377	0.0267	-0.1296	0.0106	-0.0001	0.2874
PT6		0.3390	-0.0382	0.1843	-0.0149	0.0002	-0.4096
		-0.3402	0.0383	-0.1848	0.0149	-0.0002	0.4110
		0.2383	-0.0280	0.1345	-0.0107	0.0001	-0.2998
		-0.0829	0.0094	-0.0446	0.0035	0.0000	0.1000
	2.261E-02	-0.0729	0.0755	0.0511	-0.0035	-0.0001	0.0170
		0.2460	-0.2478	-0.1635	0.0097	0.0006	-0.0628
FTIT		-0.3587	0.3557	0.2311	-0.0124	-0.0011	0.0958
		0.3611	-0.3572	-0.2316	0.0122	0.0012	-0.0971
		-0.2682	0.2619	0.1675	-0.0080	-0.0010	0.0747
		0.0927	-0.0880	-0.0547	0.0020	0.0004	-0.0277

5.2.4 Evaluation of the WSSR Statistic

In order to fully characterize the performance of any FDI algorithm, one clearly needs to consider a statistical characterization of the decision quantities used in the decision process. The algorithm of ref. [3] makes extensive use of the weighted sum of squared residuals (WSSR) statistic which takes the form,

$$d_{KF} = \sum_{j=0}^{N-1} \nu^T(k+j)\ W^{-1}\ \nu(k+j) \tag{5-36}$$

To evaluate the WSSR statistic we note that for statistics with greater than 15 degrees of freedom, the actual probability density function (which is a chi-squared function) is well characterized by its first two moments and, hence, can be considered approximately Gaussian. In the algorithm of ref [3] we have 5 residuals times 3 samples (15) degrees of freedom. The mean and covariance of d_{KF} (which involve 1st, 2nd, and 4th moments of $\nu(k)$) are given by:

$$E[d_{KF}] = N \cdot Tr[W^{-1}\ C_\nu{}^o] \tag{5-37}$$

where $C_\nu{}^o$ is the steady state covariance matrix of the approximate Gaussian density for the residual sequence, and,

$$\begin{aligned} Var\ [d_{KF}] = \sum_{k=0}^{N-1} \sum_{j=0}^{N-1} \{\ & 2 \cdot Tr[W^{-1}\ C^{j-k}\ W^{-1}\ C^{k-j}] \\ & + Tr\ [W^{-1}\ C^{j-k}] \cdot Tr\ [W^{-1}\ C^{k-j}] - Tr^2\ [W^{-1}\ C^o]\ \} \end{aligned} \tag{5-38}$$

where $C^i = E[\nu(k)\ \nu^T(k+i)]$ and is computed from Eq. 5-31.

Using the statistics of each WSSR quantity the results of Section 4 and Appendix A could be applied. Various metrics which indicate the distinguishability of the various failures modes could be computed and compared to similar metrics for other FDI algorithms.

In summary, the KF we have designed (and potentially that of ref. [3]) appears to be fairly robust to the relatively substantial model errors defined in [11]. If these were the only errors present, acceptable FDI performance would be expected. However, the full envelope algorithm must be able to predict base points as well as the system matrices which govern behavior of the engine near these points. We have shown that the ability to detect sensor failures is highly sensitive to base point uncertainty. Finally, we have developed equations for the mean and covariance of a WSSR statistic in the presence of model uncertainty and noise. These quantities can then be used to evaluate the FDI algorithm in more detail by computation of the distance metrics of Section 4.

SECTION 6

CONTRIBUTIONS AND RECOMMENDATIONS

The major contributions of this work are listed below.

1. A statistical characterization of distinguishability of failures has allowed us to evaluate FDI performance in the presence of model uncertainty and noise. Furthermore, optimization of distinguishability metrics allow us to choose the best linear relations among measurable variables for use in FDI. These relationships (called parity checks) are used to form residual signals whose behavior under different failure modes is maximally distinguishable. Unlike residuals formed in a centralized manner (e.g., using a Kalman filter based on a nominal model), the optimized parity check approach can select only parts of the system model to form residuals and thus represents a decentralized approach to FDI. Only the most well known parts of the system model are used for defining residuals. Trade-offs can then be made which incorporate notions of model uncertainty and failure distinguishability.

2. An FDI design methodology which utilizes our theoretical results has been defined. Rather than defining a "canned" design rule, which tends to obscure important design considerations, we have formulated the goals and design choices which the FDI designer must make and have identified tools which may be used in the design process.

3. The evaluation of system design issues (such as sensor configuration and model accuracy) in terms of FDI performance can be made without complete design of an FDI algorithm. The use of distinguishability metrics applied to

the measurements themselves provide lower bounds on the error probabilities of any FDI system which has access to the same measurements.

4. The evaluation of alternative FDI schemes is made possible by applying distinguishability metrics to the sufficient statistics of the algorithm. In the case of [11], this involves computing the distribution of weighted sums of squares of Kalman filter residuals. We have shown that performance is, in general, highly dependent on the choice of weights as well as Kalman Gain, and for [11] in particular, the uncertainty in linearization point plays a major role.

5. Application of our design techniques to a reduced order F-100 engine model at a single operating point has demonstrated the feasibility and usefulness of this approach. However, the highly nonlinear nature of the jet engine (both during transients and during operating point changes) produces a new issue; that of robust-adaptive FDI. Here, "adaptive" refers to the scheduling of residual generation and decisionmaking parameters with different operating conditions and during transients. The metrics developed in this project provide a useful starting point for work in this area since they allow quantitative assessment in determining the boundaries between adaptivity and robustness. That is, we can begin to answer questions such as, how good is a given linear model over a particular "parameter scheduling region", when must we change to another model, and when must nonlinear models be considered?

6.1 RECOMMENDATIONS FOR FURTHER WORK

A number of theoretical and applications oriented extensions are suggested here for future work.

1. First, the application and extension of our results to a highly nonlinear system such as the jet engine can serve to both demonstrate the

usefulness of these concepts and extend the theory and methodologies in directions which address the application specific issues in detail. For the jet engine problem, we believe that the issue of adaptivity versus robustness is of primary importance. Clearly some amount of adaptation is desirable since large differences in dynamic relationships exist between engines as well as in a single engine over its operational lifetime. A "closed loop" adaptive technique which operates on the same time scale as the FDI algorithm is not sufficient since adaptation to failures is possible. Parameter scheduling techniques should be used to transition between sets of reduced order models if one can determine the important parameters on which to schedule, and when scheduling is necessary.

In this regard, the results of this project can be applied in several interesting ways. First, in determining the validity regions of reduced order models one may consider forming sets of models through an identification algorithm operating over increasing regions in some space of parameters. As this region grows one would expect (particularly in the case of parameters being control variables and reduced order models being linear approximations to a nonlinear system) that modeling error would also grow. Quantifying the modeling error is, then, all which is needed for evaluating each model in terms of FDI performance. The largest region in parameter space which maintains a desired level of FDI performance can then be selected and the validity region for the corresponding model determined.

Alternatively one could define metrics which measure distances between the reduced order models themselves, or between reduced order models and a truth model. As in our results, these metrics would include the effects of model uncertainty in the distance definition. Clearly, a robust model is one

which is close (in terms of a metric to be defined) to the truth model even for large changes in the parameters of the reduced model. Finally, when linear models are used, uncertainty in the linearization point must be considered as a substantial source of error. That is, for stable systems, the error in predicting the steady state impulse response must be determined and included in the relevant metrics for determining model robustness.

Finally, a functional breakdown of a full envelope algorithm is presented in Fig. 6-1.

2. A second useful application to which our results may be readily applied is in the area of system design and analysis. The dependency of over-all system reliability, life cycle costs, and "maintenance down-time" on performnce (reliability) of the FDI algorithm is well known. Tradeoffs between sensor cost, weight, duplication and FDI performance can now be made without deriving an FDI algorithm for each option, and the impact on overall system issues determined. Ultimately, one might conceive of an "expert" FDI design aid which can answer relatively high level design and maintenance questions by accessing the detailed analysis provided by this and other efforts.

3. Within the FDI algorithm itself, some important issues in robust decisionmaking (i.e., how to design and evaluate the elements of the decision process) can be addressed with the results of this project. In several application areas (such as flight control) the modeling of dynamic relationships is in a fairly mature state. Thus, in the fundamental FDI decomposition (Fig. 2-1) the design of residual generation or parity relations is straight-forward (and, note, in many cases, nonlinear). In such systems, optimized parity checks may be of less interest. However, we can achieve

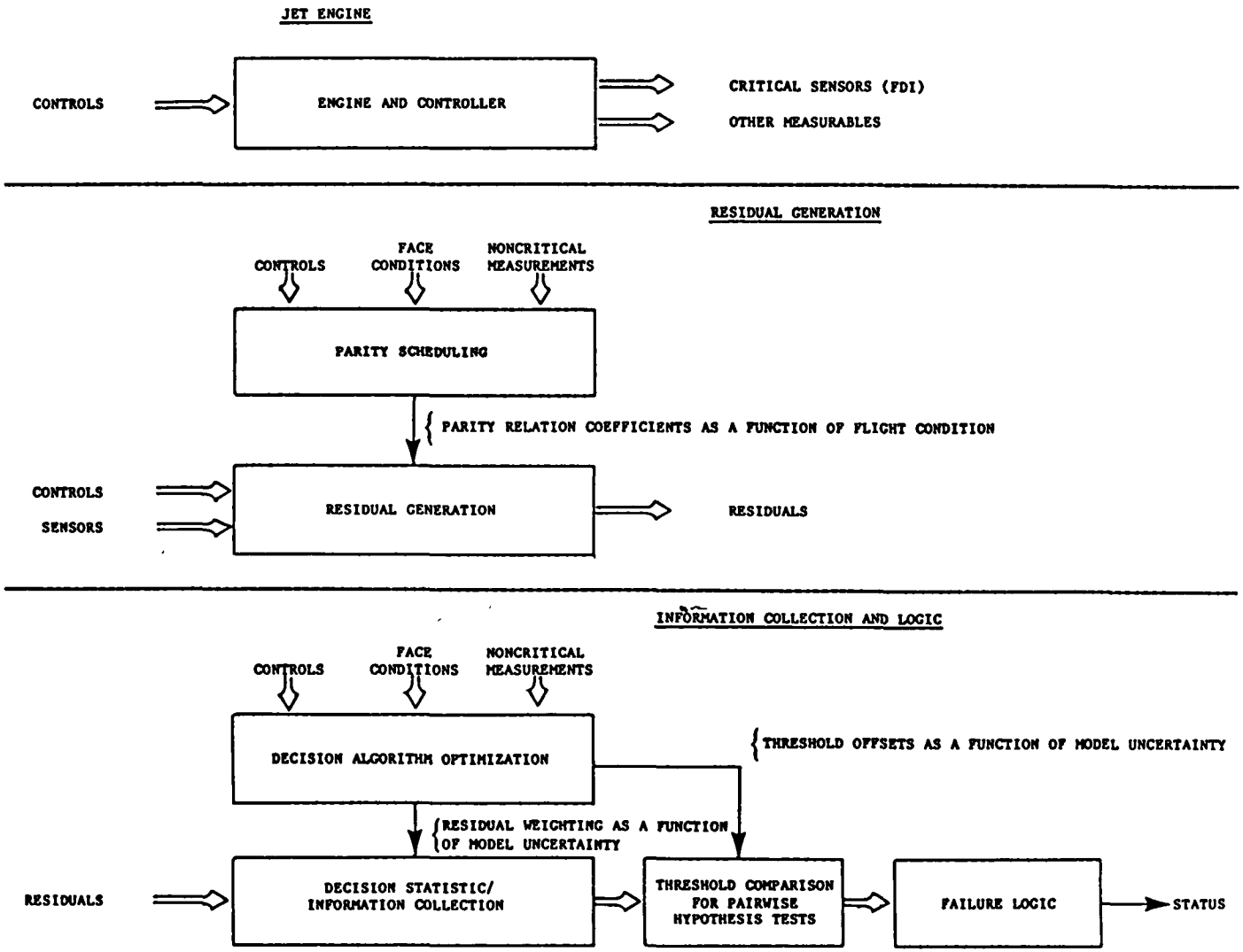


Figure 6-1. Functional Breakdown of Full Envelope Jet Engine Sensor FDI

robustness in the decisionmaking process (as discussed in Section 3) by the selective use of the known redundancy relationships. The metrics we have used could be applied to solve this "selection" problem.

In the case when knowledge about modeling uncertainty is available but varies over the same time scale as the decision algorithm, metrics provide a starting point for determining parameters of decision algorithms which might "adapt" to changing distinguishability conditions. Ultimately, in the case of changing distinguishability conditions one would like to develop a sequential testing procedure which, in effect, waits for favorable conditions before decisions are made. The relevant parameters of such a test would then be optimized at each time step.

Finally, in many cases, a decision algorithm is useful only if its performance is guaranteed for a large class of failures (e.g., all bias failures larger than a minimal failure magnitude). In some instances (e.g., tests for bias failures vs. no bias), this property is a direct result of the direct application of simple decision rules. However, it is not guaranteed in general (e.g., tests which must distinguish between two bias failure vectors, each of which has a minimal failure magnitude). Preliminary results in this regard suggest, roughly speaking, that a transformation of residuals which tends to orthogonalize the failure directions will allow decision rules to be designed for "minimal" failures with performance guaranteed for all failures which are "larger" than minimal.

4. Finally, we note than in Appendix A many connections to other technical areas (time series analysis, pattern recognition, reduced order modeling) were made. A complete clarification or unification of these areas

in terms of the goals of FDI algorithms is considered useful in obtaining better insight, and in deriving alternative concepts which may be useful for further research into robust FDI.

REFERENCES

1. Willsky, A.S., "A Survey of Design Methods for Failure Detection in Dynamic Systems," Automatica, Vol. 12, 1976, pp. 601-611; also in AGARDograph No. 224 on "Integrity in Electronic Flight Control Systems."

2. Willsky, A.S. "Failure Detection in Dynamic Systems," AGARD Lecture Series No. 109 on Fault Tolerance Design and Redundancy Management Techniques, Athens, Rome, and London, Oct. 1980.

3. Beattie, E.C., et al, "Sensor Failure Detection System Final Report, "NASA CR-165515, PWA-5756-17, NASA-Lewis Research Center Contract NAS3-22481, August 1981.

4. Chow, E.Y. and A.S. Willsky, "Issues in the Development of a General Design Algorithm for Reliable Failure Detection," Proc. of IEEE Conference on Decision and Control, Albuquerque, New Mexico, December 1980.

5. Lou, X.-C, A.S. Willsky and G.C. Verghese, "Failure Detection with Uncertain Models," preprint Sept. 1982, invited paper for 1983 American Control Conference, June 1983.

6. Deckert, J.C., M.N. Desai, J.J. Deyst, and A.S. Willsky, "F8-DFBW Sensor Failure Identification Using Analytic Redundancy," IEEE Trans. Automatic Control, Vol. AC-22, No. 5, Oct. 1977, pp. 795-803.

7. Willsky, A.S., E.Y. Chow, S.B. Gershwin, C.S. Greene, P.K. Houpt, and A.L. Kurkjian, "Dynamic Model-Based Techniques for the Detection of Incidents on Freeways," IEEE Trans. Automatic Control, Vol. AC-25, No. 3, June 1980, pp. 347-360.

8. Deckert, J.C., J.L. Fisher, D.B. Laning, and A. Ray, "A Signal Validation Methodology for Nuclear Power Plants," Proc. 1982 American Control Conference, Arlington, Virginia, June 1982, pp. 113-120.

9. Willsky, A.S. and H.L. Jones, "A Generalized Likelihood Ratio Approach to the Detection and Estimation of Jumps in Linear Systems," IEEE Trans. Automatic Control, Vol. AC-21, Feb. 1976, pp. 108-112.

10. Beard, R.V., "Failure Accommodation in Linear Systems Through Self-Reorganization," Report MVT-71-1, Man Vehicle Lab, MIT, Cambridge, Massachusetts, Feb., 1971.

11. Beattie, E.C., et al, "Sensor Detection for Jet Engines Final Report," NAS3-23282, NASA Lewis Research Center, PWA 5891-18, May 1983.

12. Pattipati, Krishna, and Alan S. Willsky, "A Statistical Distance Based Methodology for Robust Parity Generation and FDI System Evaluation," ALPHATECH Technical Memo Number TM-137, ALPHATECH, Inc., Burlington, MA, May 1984.

13. Pattipati, Krishna R., Alan S. Willsky, James C. Deckert, John S. Eterno, and Jerold Weiss, "A Design Methodology for Robust Failure Detection and Isolation," 1984 American Control Conference, San Diego, CA, June 6-8, 1984.

14. Van Trees, Harry L., Detection Estimation and Modulation Theory, Part I, Wiley & Sons, Inc., NY, 1968.

15. Kazakos, D., "On Resolution and Exponential Discrimination Between Gaussian Stationary Vector Processes and Dynamic Models," IEEE Transactions on Automatic Control, Vol. 25, No. 2, pp. 294-296.

16. Lee, C.C., and J.D. Thomas, "A Modified Sequential Detection Procedure," IEEE Transactions on Information Theory, Vol. IT-30, No. 1, June, 1984.

17. Kay, J.M., and S.L. Marple, Jr., "Spectrum Analysis - A Modern Perspective," Proc. IEEE, Vol. 69, No. 11, Nov. 1981, pp. 1380-1419.

18. Jones, H.L., "Failure Detection in Linear Systems," Ph.D. Thesis, Department of Aero. and Astro., MIT, Cambridge, MA, Sept. 1973.

19. Lou, X.-C., "A System Failure Detection Method: The Failure Projection Method," S.M. Thesis, MIT Dept. of Elec. Eng. and Comp. Sci., Rept. LIDS-TH-1203, MIT Lab for Inf. and Dec. Sys., June 1982.

APPENDIX A

TM-137

A STATISTICAL DISTANCE BASED METHODOLOGY FOR ROBUST PARITY GENERATION AND FDI SYSTEM EVALUATION

By

Krishna Pattipati
Alan S. Willsky

May 1984

Contract Number NAS3-24078

ABSTRACT

A decentralized failure detection and isolation (FDI) methodology, which is robust with respect to model uncertainties and noise, is presented. Redundancy metrics are developed, and optimization problems are posed for the choices of robust parity relations. Closed-form solutions for some special failure cases are given. Connections are drawn with other disciplines, and the use of the metrics to evaluate alternative FDI schemes is discussed.

CONTENTS

CONTENTS (continued)

FIGURES

TABLES

ACKNOWLEDGEMENT

We acknowledge, with thanks, the technical discussions we had with ALPHATECH staff members, Jerold L. Weiss, James C. Deckert and John S. Eterno. The suggestions of Dr. Walter Merrill, NASA-Lewis Contract Monitor, during project review meetings helped clarify several key aspects of our FDI design approach.

A-1. INTRODUCTION

A-1.1 BACKGROUND AND MOTIVATION

Sensor failure detection and isolation (FDI) deals with the problem of detecting deviations from normal behavior (which we will call "failure modes") in a specified sensor complement and isolating the particular instrument that has failed. In recent years, numerous approaches have been developed to perform FDI in linear, dynamic and stochastic systems. These methods include the voting schemes [1],[4], the generalized likelihood ratio (GLR) technique [3], [4], the multiple model (MM) approach [5],[6] and the detection filter scheme [7],[8],[9]. Willsky [10] has provided a comprehensive survey of the various FDI methods and a discussion of the strengths and weaknesses of each method.

In general, all FDI methods use, either implicitly or explicitly, redundancy relations among the measured system variables to generate residuals (or generalized parity checks) and make failure decisions. In addition, all FDI methods can be conceptualized as consisting of three separate, clearly identifiable stages: 1) residual (parity) generation, 2) information collection and 3) decisionmaking (see Fig. A-1-1). Essentially, the parity checks are a set of signals that should be near zero when the system is operating normally, and that will deviate from zero in characteristic ways when particular failures occur. The residual (parity) generation process can be of varying complexity for different FDI systems. For example, in a voting scheme parity checks are simply the differences of outputs from identical sensors, whereas in the GLR system, the parity checks are generated by the more complex Kalman filter (KF).

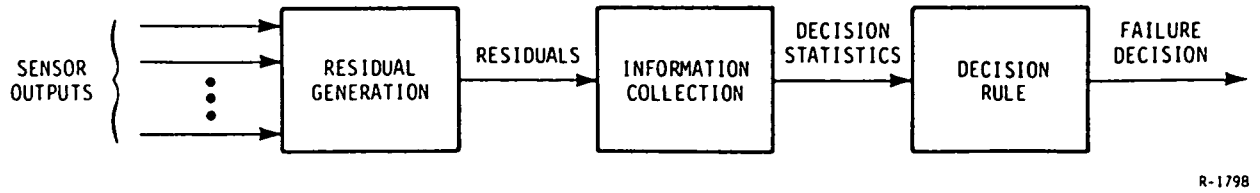


Figure A-1-1. Three Stage Structure of the FDI Process.

In the information collection phase, the parity checks are accumulated over time (typically over a moving time window) for the purpose of establishing the presence of failures. This phase may produce one or several scalar decision statistics, which are then compared with a threshold in the decision phase to make failure decisions. The scalar statistics in most FDI systems are probabilities or likelihood ratios for distinguishing among a set of hypotheses on residual behavior (where each such hypothesis corresponds to a particular failure mode). These statistics are typically generated in one of two ways ... via coherent or incoherent processing ... as shown in Fig. A-1-2.

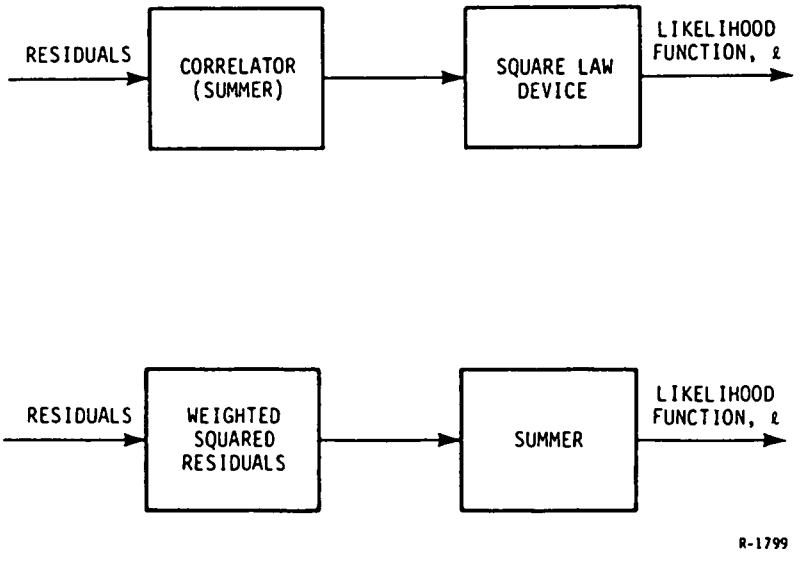


Figure A-1-2. Information Collection Methods

In the coherent processing, one correlates the observed residual sequence with a specified function of time and then squares the resulting quantity. Here, what one, in essence, assumes is that one knows the sign history of each residual component and any important phase relationship among residual components under any particular failure mode. The prototypical example of this is when a failure manifests itself as a slowly-varying bias or drift in a particular residual component. Coherent processing in this case amounts to weighted summing of measurements followed by squaring.

In the incoherent processing, one first computes particular weighted squared sums of the residuals and then sums the quantities over time. In this case, one is essentially assuming that a particular linear combination of the residual components is large at every point in time but that the temporal variation of this linear combination is unknown. Such a structure is of use, for example, in detecting noise variance changes.

The tradeoff between coherent and incoherent processing is relatively clear. In the coherent case, we are matching much more closely a particular temporal variation and therefore will reject a larger class of other variations. Thus, if we really know this variation (i.e., the sign history of residuals under a particular failure mode), one would expect a superior detection-false alarm performance using the coherent processor. If, on the other hand, the actual variation pattern under failed conditions is significantly different from the assumed one, the coherent processor will not perform nearly as well, and one may have to settle for the more conservative detection-false alarm capabilities of the incoherent processor.

The decision rule in an FDI algorithm normally involves comparing the decision statistics with one or more thresholds. The thresholds are determined to optimize a detection criterion such as the Bayes' risk, probability of error, etc.

Since analytic model-based redundancy is the key to any FDI system, the robustness of any such system depends on the reliability of the redundancy (parity) relations that are used, given the inevitable presence of model uncertainties, nonlinearities, and noise. That is, the problem of robust failure detection and isolation is concerned with designing the parity generation and information collection phases of the FDI system so that the resulting decision statistics are "maximally sensitive" to the failure modes and "minimally sensitive" to model uncertainties and noise. This perspective was at the heart of the F-8 sensor FDI project [11], and the more recent theoretical work reported in [12], [13].

A-1.2 FDI DESIGN PHILOSOPHY

The design of a practical FDI system requires consideration of several issues:

1. Performance of the detection system as measured by the detection time and detection accuracy (correct detections and false alarms).

2. Robustness, i.e., relative insensitivity of the FDI system performance to parameter variations and modeling errors or uncertainties.

3. Computational complexity as measured by storage requirements and computation time.

One could imagine formulating an overall robust FDI system design problem which uses detailed dynamical models and which accounts for model uncertainties. This alternative, however, is of conceptual value only as such a problem

is far more complex than can be solved practically as well as being more complex than is necessary to obtain a practical design. A second approach - which, in some sense, is the default approach used in most top-down attempts at FDI design - is to synthesize an FDI system based on a nominal model, neglecting model uncertainties. One then evaluates FDI performance in the presence of model uncertainties and then modifies the design to improve performance e.g., one adds hypotheses corresponding to model errors in order to alert the FDI system to the fact that such uncertainties must be distinguished from failures. This more or less trial and error approach has obvious drawbacks in terms of providing very little in the way of a concrete methodology and, more importantly, the designs obtained in this manner are often unwieldy. This is because various "bells and whistles" are generally added to alert the FDI system to the fact that part of the system may be in error. Consequently, it is not a simple manner to check all possible "what-if" scenarios in order to troubleshoot possible robustness problems.

The third approach to robust FDI system design is motivated by a desire to overcome the limitations just mentioned by taking uncertainty into account at the start in order to obtain as simple an FDI system as possible and that makes optimum use of the best information available (here simplicity is important not just for computational reasons but, more importantly, because simply structured systems are far easier to troubleshoot in order to pinpoint potential weakpoints and performance-limiting issues). This approach, which can be thought of as the decentralization of the FDI process, is based on the concept of identifying only those parts of the system model that are known with greater certainty and then design the system based primarily on the certain parts. The key idea here is that in such a design one can focus information

for a particular failure mode using only those parts of the model which provide the best "failure signature-to-model uncertainty ratio" for this failure. This decentralized approach is in marked constrast to a centralized FDI design based on one complete model, say using a Kalman filter, in which failure information and all sources of model uncertainty are diffused through the entire set of residuals. Consequently, when information collection (as in Fig. A-1-2) is performed for such an approach, one will, in general, be mixing together high quality and low quality information and the result may be a loss in robustness. Another major advantage of the decentralized approach is that it results in an FDI system in which the computations required to detect and isolate individual failures are typically far simpler than in a centralized approach. Furthermore the decentralized structure allows greater flexibility in designing an architecture for its implementation than a centralized FDI design. Finally, because parity relations are explicitly chosen for particular purposes (e.g., detecting a particular failure), detection logic is extremely simple.

For the reasons just outlined we have taken as the basis of our approach the design of robust residual generation and information collection systems. This approach was, in essence, used in the successful F-8 sensor FDI project [11] and in the theoretical work by Chow and Willsky [12] and Lou, Willsky, and Verghese [13] which formed the foundation upon which the work in the present project has been built.

A-1.3 OVERVIEW OF THE DESIGN APPROACH

In the next two sections, we describe the major analytical constructs that are needed in our approach. We preface that development here with a brief overview of the major concepts behind the approach. The first part of

our design methodology is to identify those portions of the system dynamics that are known with the most certainty, as the use of parity checks based on these portions of the dynamics will be of great value in minimizing false alarms. Thus the first problem is to obtain a rank-ordered list of parity relations, where the ranking is in terms of some "robustness metric" that quantifies how close to zero each parity check is under normal conditions given the presence of model errors and noise.

The second problem is coverage - that is, assessing the ability of relations identified in the first step to detect a specified set of failure modes (each of which is also modeled with an allowance for errors in the model used) and the possible augmentation of this set with additional relations that are useful in detecting particular failure modes not well-covered by the initial group of relations. Let us note two aspects of this problem. The first is that it requires a second metric that measures the ability of a particular relation to distinguish between normal conditions and a particular failure mode (i.e., to give decidely larger values under the particular failure than under normal conditions) given the presence of model error and noise. Again this metric can be used to rank-order relations with respect to their usefulness for detecting particular failures. The second point is that the relations added at this stage are less reliable than the first ones obtained, in that they may have larger values when no failure has occurred. They may be needed, however, to achieve the desired coverage (i.e., to achieve a specified probability of detection for <u>all</u> failures). Note also that the metric used should provide a guide to the minimum magnitude of a particular failure that can be reliably detected (i.e., to achieve a specified probability of false alarm).

The final step in the design is concerned with the problem of distinguishability: given that a failure has occurred, can we determine which failure mode has occurred. Here again we need a metric that measures the ability of a parity relation to distinguish a particular failure mode from an alternative set of possible events corresponding to one or more of the other failure modes. Note again that if additional relations ar needed, they will be inherently less reliable under normal conditions. However, at this stage we can, if necessary, avoid the impact such relations would have on false alarm rate by using a two-level structure. Specifically, the relations determined in the first two stages are sufficient for detection of all failures, but may not be able to isolate all of them. Thus, one could use these relations to detect and trigger the use of additional relations for isolation only.

For the most part we have discussed the residual generation phase of the FDI system. However, the same ideas are also of value in designing the information collection phase, i.e., in determining the data length required to achieve desired performance levels and in specifying the details of how successive residuals are accumulated. In the next two sections, we describe the metrics we have considered and indicate how they are used in the design of robust FDI systems. We also provide a brief discussion of the interpretations of parity relations as reduced-order models and the use of our metrics to assess the robustness of other FDI systems.

A-2. ROBUST PARITY (RESIDUAL) GENERATION METHODOLOGY

As we stated in Section 1, the key to robust FDI is the "pulling apart" of the most reliable of the available redundant relations among the measured variables. In this section, we provide a precise definition and a unified view of redundancy, and develop a methodology for generating robust redundancy relations in the presence of model uncertainties. We will begin our discussion of redundancy with static perfectly known systems with no noise.

A-2.1 REDUNDANCY AND PARITY RELATIONS: STATIC PERFECTLY KNOWN SYSTEMS WITH NO MEASUREMENT NOISE [14],[15]

Suppose that the redundant measurements can be modeled by the measurement equation:

$$y = Cx \tag{A-2-1}$$

where y is an m vector of measurements, C is the mxn measurement matrix of rank r and x is the n vector of state variables. In static systems, one can interpret redundancy relations in four equivalent ways as follows.

A-2.1.1 Projection Matrix Interpretation

The unique minimum norm, least squares solution of Eq. A-2-1 is given by

$$\hat{x} = C^{\dagger}y \tag{A-2-2}$$

where $C^{\dagger}$ is the generalized (Moore-Penrose) inverse of C. We can view the estimate in Eq. A-2-2 as the best linear unbiased estimate with infinite prior covariance for x. The residual vector, μ is given by

$$\mu = y - C\hat{x} = (I_m - CC^{\dagger})y = P_c y \qquad \text{(A-2-3)}$$

where P_c is the mxm projection matrix, with the following properties:

1. P_c spans an (unobservable) subspace that is orthogonal to the (observable) subspace spanned by the rows of C. That is,

$$P_c.C = (I_m - CC^{\dagger})C = C - CC^{\dagger}C = 0 \text{ (From Moore-Penrose inverse property)} . \qquad \text{(A-2-4)}$$

Also, the residuals satisfy the relation:

$$\mu = P_c\, y = 0 \rightarrow \hat{x} \text{ matches the measurements y exactly .}$$

2. Rank of P_c = m - rank of (C) = m - r.

3. The rows of the projection matrix, P_c provide m-r=t <u>independent, algebraic relationships</u> among the m measured variables. Since these relationships are algebraic, we call them <u>parity checks of order 0</u> (or 0-th order parity system). The t independent rows of P_c are the t <u>parity vectors</u> of the 0th order parity system. Note that the parameter synthesis approach of [16] is a 0th order parity system.

4. The projection matrix P_c is idempotent and symmetric, i.e.,

$$P_c = P_c^{\,2}$$

$$P_c = P_c^{\,T}$$

A-2.1.2 <u>Singular Vector Interpretation</u>

Consider the singular value decomposition (SVD) of the measurement matrix C

$$C = U\ \Sigma\ V^T \qquad \text{(A-2-5)}$$

where

U = mxm matrix of left singular vectors

Σ = mxn matrix of singular values

V = nxn matrix of right singular vectors

The matrices U and V are orthonormal with the following properties:

$$UU^T = U^TU = I_m$$

$$VV^T = V^TV = I_n \qquad \text{(A-2-6)}$$

When rank [C] = r, the SVD (Eq. A-2-5) can be rewritten explicitly as:

$$C = m\ \overset{t}{[\ U_1}\ :\ \overset{r}{U_2\ ]}\ \begin{matrix} t \\ r \end{matrix} \left[\begin{array}{c|c} 0 & 0 \\ \hline 0 & \Sigma_2 \end{array}\right] \begin{matrix} \overset{n-r}{} \\ \end{matrix} \left[\begin{array}{c} V_1^T \\ \hline V_2^T \end{array}\right] \begin{matrix} \overset{n}{} \\ \begin{matrix} n-r \\ r \end{matrix} \end{matrix}$$

$$= U_2\ \Sigma_2\ V_2^T \quad ; \quad \Sigma_2 > 0 \qquad \text{(A-2-7)}$$

where

$$U_1^TU_1 = I_t \quad ; \quad U_1^TU_2 = 0_{t,r}$$

$$U_2^TU_2 = I_r \quad ; \quad U_2^TU_1 = 0_{r,t} \qquad \text{(A-2-8)}$$

Using Eq. A-2-8,

$$U_1^TC = 0 \qquad \text{(A-2-9)}$$

Eq. A-2-9 says that the t left singular vectors corresponding to zero singular values of C are the t independent parity relations:

$$\mu = U_1^Ty = U_1^TCx = 0 \qquad \text{(A-2-10)}$$

where $\mu = [\mu_1 \mu_2 \cdots \mu_t]^T$. The columns of U_1 are the t parity vectors $u_1, u_2, \cdots, u_t$; we call U_1 the parity check matrix. A remarkable property of the parity vectors u_i is that they form an orthonormal set of vectors in the measurement space, i.e., they are as separate as possible.

$$u_i^T u_j = 0 \text{ for all } i \neq j \quad . \tag{A-2-11}$$

Remark

The projection matrix, P_c and the t left singular vectors, U_1 are related by

$$P_c = U_1 U_1^T \ .$$

This can be seen as follows:

$$P_c = I_m - CC^\dagger = U_1 U_1^T + U_2 U_2^T - U_2 \Sigma_2 V_2^T V_2 \Sigma_2^{-1} U_2^T = U_1 U_1^T$$

since

$$C^\dagger = V_2 \Sigma_2^{-1} U_2^T \text{ and } U_2^T U_2 = I_r \quad .$$

Example 2.1: Consider a dual redundant sensor system where signal of sensor 2 is related to the signal of sensor 1 by a scale factor α. That is,

$$y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 \\ \alpha \end{bmatrix} x = C x \ ; \ \text{Rank } [C] = r = 1 \quad .$$

a. Projection Matrix Approach

$$C^\dagger = \begin{bmatrix} \frac{1}{1+\alpha^2} & \frac{\alpha}{1+\alpha^2} \end{bmatrix}$$

$$P_c = I - CC^\dagger = \begin{bmatrix} \frac{\alpha^2}{1+\alpha^2} & \frac{-\alpha}{1+\alpha^2} \\ \frac{-\alpha}{1+\alpha^2} & \frac{1}{1+\alpha^2} \end{bmatrix} \ , \ \text{Rank } [P] = m-r = 1 \quad .$$

Thus, P_c defines a single parity check equation

$$\mu_1 = \alpha y_1 - y_2 = 0 \quad .$$

b. SVD Approach

The SVD of C is

$$C = \begin{bmatrix} 1 \\ \alpha \end{bmatrix} = \underbrace{\begin{bmatrix} \frac{\alpha}{\sqrt{1+\alpha^2}} & \vdots & \frac{1}{\sqrt{1+\alpha^2}} \\ \\ \frac{-1}{\sqrt{1+\alpha^2}} & \vdots & \frac{\alpha}{\sqrt{1+\alpha^2}} \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} \frac{0}{\sqrt{1+\alpha^2}} \end{bmatrix}}_{\Sigma} \cdot \underbrace{1}_{V^T} \quad .$$

The left singular vector corresponding to the zero singular value is:

$$u_1 = \begin{bmatrix} \frac{\alpha}{\sqrt{1+\alpha^2}} \\ \\ \frac{-1}{\sqrt{1+\alpha^2}} \end{bmatrix} \quad .$$

The parity check equation $u_1^T y = 0$ becomes:

$$\mu_1 = \alpha y_1 - y_2 = 0 \quad .$$

It is easy to see that $P_c = u_1 u_1^T$.

A-2.1.3 Eigenvector Interpretation

Another interpretation of parity vectors, that will be useful in later sections, is that they are the eigenvectors of CC^T. This can be seen by considering CC^T in terms of the SVD as:

$$
\begin{aligned}
CC^T &= U \Sigma V^T V \Sigma U^T = U \Sigma^2 U^T \\
\rightarrow CC^T U &= U \Sigma^2 \\
CC^T u_i &= \sigma_i^2 u_i \quad ; \quad i = 1,2,\ldots,m \quad .
\end{aligned}
$$

Thus, the left singular vectors u_i are the eigenvectors of CC^T and the parity vectors are the eigenvectors of CC^T corresponding to zero eigenvalues. Since $tr(CC^T)$ is a measure of energy in the measured signals, parity vectors can be thought of as the directions of least energy. This observation leads us to a stochastic interpretation in terms of minimum entropy, whenever the state x is random. We explore this interpretation next.

A-2.1.4 Minimum Entropy Interpretation

Entropy is a statistical measure of uncertainty [17]. As we will now argue, the entropy concept can be used as a suitable criterion for parity selection. In particular, a parity check can be thought of intuitively as a variable which, under unfailed conditions, takes on as close to a specified deterministic value (typically zero) as possible. The less randomness there is in a parity check, the more useful it is for detecting failures. Thus, if we view entropy as a measure of uncertainty, a meaningful parity selection criterion is to choose parity relations which minimize the entropy (uncertainty) of y. Since this criterion is equivalent to minimizing variations of μ in the transformed domain (i.e., parity space), it is reasonable to expect that the transformed measurements (i.e., parity relations) will have clustering

properties (i.e., in example 2.1, parity check $\mu = y_2 - \alpha y_1 = 0$, is clustered around the origin).

To prove the minimum entropy interpretation of parity relations and to provide connections to the eigenvector interpretation, assume that x is Gaussian with zero mean and covariance, Σ , i.e., $x \sim N(0,\Sigma)$. We assume that $\Sigma > 0$. The basic idea of minimum entropy approach is to select a parity transformation

$$\mu = G^T y \qquad \text{(A-2-12)}$$

such that the entropy of parity relations, μ is minimized. In Eq. A-2-12, y is an m vector, μ is a t vector of desired number of parity relations and G is an m by t parity check matrix. To avoid trivial solutions G=0, we impose the orthonormality constraint on G:

$$G^T G = I_t \quad . \qquad \text{(A-2-13)}$$

Note that the covariance of y is $C\Sigma C^T$, and this will be invertible if and only if C has full row-rank. In this case there are no <u>perfect</u> parity checks as discussed in the preceding sections (where $\mu \equiv 0$ under unfailed conditions), but what we wish to choose are parity checks so that the uncertainty in μ is as small as possible. If C has rows that are linearly dependent, then there are perfect parity checks, as discussed in the preceding subsections, corresponding to the eigenvectors associated with the zero eigenvalue of $C\Sigma C^T$. If one wishes to construct additional parity checks, then one again must settle for some level of uncertainty. These can be constructed using the procedure we now describe but first removing the dependent rows of C, thereby producing a matrix C with the same rank as C. For these reasons, in the remainder of this section, we assume without loss of generality that the rows of C are linearly independent.

The entropy of μ is given by

$$H_\mu(t) = - \int_\mu p(\mu) \ln p(\mu) \, d\mu \qquad \text{(A-2-14)}$$

where we have indicated the explicit dependence of H_μ on the number of parity relations, t. Since μ is Gaussian with zero mean and covariance matrix $G^T C \Sigma C^T G$,

$$H_\mu(t) = \frac{t}{2} \ln(2\pi e) + \frac{1}{2} \ln \det \left(G^T C \Sigma C^T G\right) \quad . \qquad \text{(A-2-15)}$$

Thus, entropy $H_\mu(t)$ is a function of the covariance function of μ. The optimum transformation matrix G is given by the following theorem.

Theorem 2.1

The entropy function H_μ is minimized by taking the columns of G to be the t normalized eigenvectors associated with the smallest t eigenvalues of the covariance matrix of y, $C \Sigma C^T$.

Proof:

Entropy $H_\mu(t)$ in Eq. A-2-14 is minimized with respect to G by forming the augmented Lagrangian function:

$$H_\mu' = \frac{t}{2} (\ln 2\pi e) + \frac{1}{2} \ln \det \left(G^T C \Sigma C^T G\right) + \frac{1}{2} \operatorname{tr} \left[\left(G^T G - I_t\right) \cdot \Gamma\right] \qquad \text{(A-2-16)}$$

where Γ is the matrix of Lagrange multipliers. Near optimal G^* we have, for small ΔG

$$\Delta H_\mu' = \frac{1}{2} \ln \det (G^* + \Delta G)^T \, C \Sigma C^T \, (G^* + \Delta G) - \frac{1}{2} \ln \det \left(G^{*T} C \Sigma C^T G^*\right)$$

$$+ \frac{1}{2} \operatorname{tr} \left[\left((G^* + \Delta G)^T (G^* + \Delta G) - G^{*T} G^*\right) \Gamma\right] \qquad \text{(A-2-17)}$$

Using $(I + \Delta)^{-1} \simeq I - \Delta$, and $\ell n \det (I + \Delta) \simeq tr(\Delta)$, we have the first order condition:

$$\Delta H_\mu' = tr\left[\Delta G^T\{C\Sigma C^T G^* \left(G^{*T} C\Sigma C^T G^*\right)^{-1} - G^* \Gamma\}\right] = 0 \quad . \tag{A-2-18}$$

(Here, we are using the fact that $G^{*T}C\Sigma C^T G^*$ is invertible. This will be the case if C has full rank.) In order to satisfy Eq. A-2-18 regardless of ΔG, we need

$$C\Sigma C^T G^* \left(G^{*T} C\Sigma C^T G^*\right)^{-1} = G^* \Gamma \quad . \tag{A-2-19}$$

Premultiplying Eq. A-2-19 by G^{*T} and using $G^{*T}G^* = I_t$, we have $\Gamma = I_t$. Using $\Gamma = I_t$, Eq. A-2-19 can be rewritten as

$$C\Sigma C^T G^* = G^* \cdot \left(G^{*T} C\Sigma C^T G^*\right) \quad . \tag{A-2-20}$$

From Eq. A-2-20 it is clear that the columns of G^* are the eigenvectors of $C\Sigma C^T$ (in this case note that $G^*C\Sigma C^T G^{*T}$ is diagonal), and thus G* whose columns consist of t eigenvectors of $C\Sigma C^T$ is a stationary point of the entropy function. It is immediate, then, from Eq. A-2-15 that the minimum is achieved by choosing the eigenvectors corresponding to the t smallest eignevalues of $C\Sigma C^T$.

Remarks

1. The minimum entropy result we just obtained provides us with a convenient interpretation in terms of eigenvector and eigenvalue analysis of the covariance function of y: the parity vectors are the eigenvectors of the covariance function of y corresponding to the smallest eigenvalues. This observation, and the minimum entropy interpretation enables us to extend the parity space concepts to dynamic and stochastic systems. More importantly, this result says that we can work directly with the experimentally obtained

covariance data to generate the parity relations, if a model of the system is not available.

2. Since eigenvectors corresponding to the smallest eigenvalues of the covariance matrix are the directions of "near perfect" correlation, minimum entropy transformation provides the "most certain" or deterministic relationships among the measured variables, i.e., relations with least variability or relations having clustering properties. Thus, the eigenvalues associated with parity vectors are measures of their certainty (or "robustness").

3. The use of minimum entropy criterion to feature selection process in pattern recognition literature is well known [17]. The criterion is used because of the clustering effects it produces on the data, i.e., the parity check values are as close to a specified set of deterministic values as possible.

4. Note that if x is equally likely to be in any direction, i.e., $\Sigma=\alpha I$, the parity vectors correspond to eigenvectors of $CC^T\cdots$, i.e., when we have no a priori information about the relative scaling of components of x, we recover the results of the preceding subsections.

A-2.2 REDUNDANCY AND PARITY RELATIONS: STATIC PERFECTLY KNOWN SYSTEMS WITH MEASUREMENT NOISE.

Consider a measurement system of the form

$$y = Cx + v \qquad \text{(A-2-21)}$$

where x is a Gaussian state vector with zero mean and covariance, Σ; and v is a zero mean, Gaussian noise process with covariance R, which is uncorrelated with the original process, x.

As before, we seek a parity transformation matrix G to minimize the entropy of parity relations, $\mu = G^T y$ given by

$$H_\mu = \frac{t}{2} \ln (2\pi e) + \frac{1}{2} \ln \det \left[G^T (C\Sigma C^T + R)\, G \right] \quad . \qquad \text{(A-2-22)}$$

As before the entropy function H_μ is minimized by forming the transformation matrix G from the t normalized eigenvectors associated with the smallest eigenvalues of the covariance matrix, $(C\Sigma C^T+R)$.

Remarks

1. The minimum entropy orthonormal transformation G also minimizes the scatter (dispersion measure) given by

$$d_\mu^2 = \text{tr} \left(E \{ \mu\, \mu^T \} \right) = \text{tr} \left(G^T (C\Sigma C^T + R) G \right) \quad . \qquad \text{(A-2-23)}$$

2. In the statistical literature, the use of eigen analysis to characterize features is also referred to as factor analysis or principal component analysis (PCA) [18]. PCA has been used in model order reduction problems [19], system identification [20], and image processing [21].

3. A nice property of the parity relations μ is that they are component wise uncorrelated, i.e.,

$$E \{ \mu_i \mu_j \} = 0 \text{ for all } i \neq j,\ i,j = 1,2 \cdots t \quad . \qquad \text{(A-2-24)}$$

Under Gaussian assumption, the components are independent as well.

4. When measurement noise is present, the minimum entropy transformation G provides parity relations that are as orthogonal as possible to both the signal and the noise processes.

5. In pattern recognition literature [17],[22], it is a common practice to use the autocorrelation function of the measurement process y, which is also the covariance matrix of y under the zero mean assumptions on x and v. The autocorrelation function approach has the important interpretation that the parity vectors are the directions of least energy.

A-2.3 REDUNDANCY AND PARITY RELATIONS: STATIC UNCERTAIN SYSTEMS WITH MEASUREMENT NOISE

Consider a static, uncertain measurement system

$$y = c_{\ell}x + v_{\ell} \tag{A-2-27}$$

where the measurement matrix C_{ℓ} takes on values from a finite set of L models, $\ell = 1,2,\cdots,L$; v_{ℓ} is a zero mean, white-Gaussian noise process with covariance R_{ℓ} under the model hypothesis ℓ; x is an n-dimensional state vector with zero mean and covariance Σ_{ℓ} under model hypothesis ℓ. In addition, the a priori probability that the ℓ-th model is correct is, P_{ℓ}. Note that the probability density of y is a sum of L Gaussian densities:

$$p(y) = \sum_{\ell=1}^{L} P_{\ell}\, N\left(0, C_{\ell}\Sigma_{\ell}C_{\ell}^{T} + R_{\ell}\right) \quad . \tag{A-2-28}$$

Our objective is to find the minimum entropy parity transformation such that:

$$\mu = G^{T} y \tag{A-2-29}$$

subject to the normalization constraint:

$$G^{T}G = I_{t} \quad . \tag{A-2-30}$$

The entropy of μ for a given model hypothesis ℓ is

$$H_{\mu/\ell} = \frac{t}{2}\left(\ell n 2\pi e\right) + \frac{1}{2}\,\ell n \det\left(G^{T}(C_{\ell}\Sigma_{\ell}C_{\ell}^{T} + R_{\ell})G\right) \quad . \tag{A-2-31}$$

The average entropy of μ over the range of model hypotheses is the weighted sum of conditional entropies given by Eq. A-2-31. That is,

$$H_{\mu}(t) = \sum_{\ell=1}^{L} P_{\ell}\, H_{\mu/\ell} \tag{A-2-32}$$

where we have shown the explicit dependence of H_{μ} on the desired number of

parity relations t. Proceeding as in the proof of theorem 2-17, the minimum entropy parity transformation G^* satisfies the nonlinear equation:

$$G^*\Gamma = \sum_{\ell=1}^{L} P_\ell \left(C_\ell \Sigma_\ell C_\ell^T + R_\ell\right) G^* \left(G^{*T}(C_\ell \Sigma_\ell C_\ell^T + R_\ell)G^*\right)^{-1} \qquad \text{(A-2-33)}$$

where Γ is the matrix of Lagrange multipliers. Pre-multiplying Eq. (A-2-33) by G^{*T}, and noting that $G^{*T}G^* = I_t$, we have $\Gamma = I_t$. Thus, the optimal parity check matrix G^* satisfies the nonlinear matrix relationship:

$$G^* = \sum_{\ell=1}^{L} P_\ell \left(C_\ell \Sigma_\ell C_\ell^T + R_\ell\right) G^* \left(G^{*T}(C_\ell \Sigma_\ell C_\ell^T + R_\ell)G^*\right)^{-1} . \qquad \text{(A-2-34)}$$

A simple successive substitution scheme can be employed to solve for the optimal minimum entropy parity transformation matrix. At iteration i, the idea is to find $G^{(i)}$ via *

$$G^{(i)} = \sum_{\ell=1}^{L} P_\ell \left(C_\ell \Sigma_\ell C_\ell^T + R_\ell\right) G^{(i-1)} \left[G^{(i-1)T}(C_\ell \Sigma_\ell C_\ell^T + R_\ell)G^{(i-1)}\right]^{-1} \qquad \text{(A-2-35)}$$

The algorithm starts with an initial $G^{(0)}$ and iteratively updates till $\|G^{(i)} - G^{(i-1)}\|$ is small with respect to some suitable norm. However, if we approximate the density function of y (or μ), which is a sum of L Gaussian densities with a single Gaussian density having the same mean and covariance an important simplification arises. That is, we approximate

$$\sum_{\ell=1}^{L} P_\ell \, N(0, C_\ell \Sigma_\ell C_\ell^T + R_\ell) \sim N(0, \tilde{C}) \qquad \text{(A-2-36a)}$$

*Alternatively one can use conjugate gradient method to solve for G using a modified form of Eq. 2-34 as the descent direction.

where

$$\hat{C} = \sum_{\ell=1}^{L} P_\ell \left(C_\ell \Sigma_\ell C_\ell^T + R_\ell\right) \quad . \qquad \text{(A-2-36b)}$$

With this assumption, we can state Theorem 2.2.

Theorem 2.2 Assumption 2-36 implies that an optimal choice of G is the t eigenvectors of $\hat{C}$ corresponding to the smallest t eigenvalues of $\hat{C}$.

Proof: Same as Theorem 2.1.

Example 2.2

Consider the dual redundant measurement system with three uncertain measurement matrices

$$C_1 = \begin{bmatrix} 1 \\ 1.2 \end{bmatrix} \; ; \; C_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \; ; \; C_3 = \begin{bmatrix} 1 \\ .8 \end{bmatrix}$$

$$P_1 = P_2 = P_3 = \frac{1}{3} \; ; \; R_1 = R_2 = R_3 = \begin{bmatrix} .1 & 0 \\ 0 & .1 \end{bmatrix} \; ; \; \Sigma_\ell = 1, \; \ell = 1,2,3$$

Using the Gaussian sum approximation,

$$\hat{C} = \begin{bmatrix} 1.1 & 1 \\ 1 & 1.13 \end{bmatrix}$$

Minimum eigenvalue ≈ .12

Eigenvector = $[.714 \; -.7]^T$

The robust parity vector is illustrated in Fig. A-2-1.

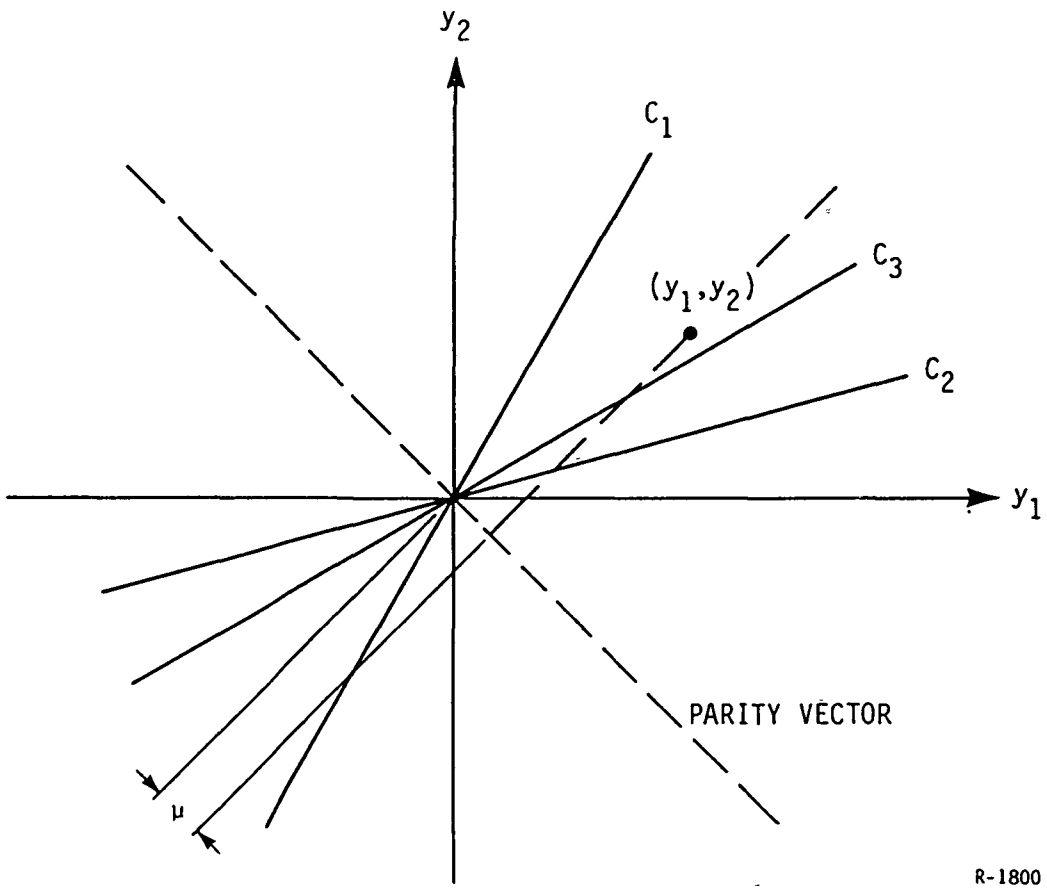


Figure A-2-1. Illustration of Robust Parity Vector in Static Uncertain Systems

A-2.4 REDUNDANCY AND PARITY RELATIONS: LINEAR DYNAMIC AND STOCHASTIC SYSTEMS WITH MODEL UNCERTAINTY

A-2.4.1 Problem Formulation

Consider a linear time-invariant, discrete-time stochastic system model

$$x(k+1) = A_\ell\, x(k) + D_\ell\, w(k) \qquad \text{(A-2-37)}$$

$$y(k) = C_\ell\, x(k) + v_\ell(k) \qquad \text{(A-2-38)}$$

where the system parameters $\{A_\ell, D_\ell, C_\ell\}$ take on values from a finite set of L models, $\ell=1,2,\cdots,L$; x is an n-dimensional state vector with zero mean and steady-state covariance matrix, Σ_ℓ under the model hypothesis ℓ ; y is the m-dimensional output vector; w is an n_w dimensional, zero-mean, white-Gaussian noise process with covariance matrix Q ; and v_ℓ is an m-dimensional, zero-mean, white-Gaussian noise process with covariance matrix R_ℓ. In addition, the a priori probability that the ℓ-th model is correct is P_ℓ .

Following [12],[13], we define a parity check of order p as a linear combination of the lagged and present values of the output over a window of size p such that the parity checks have small values if no failure occurs. Let $\mu^T(k) = [\mu_1(k), \mu_2(k), \cdots, \mu_i(k), \cdots, \mu_t(k)]$ denote a t-vector of such parity relations. Then, the parity relation $\mu_i(k)$ is related to the lagged and present values of the outputs by:

$$\mu_i(k) = g_i^T Y_p(k) \quad ; \; i=1,2,\cdots,t \qquad \text{(A-2-39)}$$

where $Y_p(k)$ and g_i, the i-th parity vector, are

$$Y_p^T(k) \triangleq [y(k-p),\, y(k-p+1),\, \cdots,\, y(k)] \qquad \text{(A-2-40a)}$$

$$g_i^T \triangleq (g_{i0},\, g_{i1},\, \cdots,\, g_{ip}) \,. \qquad \text{(A-2-40b)}$$

We also write

$$\mu(k) = G^T Y_p(k) \qquad \text{(A-2-41a)}$$

where

$$G \triangleq [g_1, g_2, \cdots, g_t] \ . \qquad \text{(A-2-41b)}$$

The window of outputs $Y_p(k)$ satisfies a linear relationship of the form:

$$Y_p(k) = M_{p\ell}\, x(k-p) + \Pi_{p\ell}\, w_p(k) + v_{p\ell}(k) \qquad \text{(A-2-42)}$$

where $w_p(k)$ and $v_{p\ell}(k)$ are lagged and present values of the process noise and measurement noise processes over a window of size p:

$$w_p(k) \triangleq [w(k-p),\ w(k-p+1),\ \cdots,\ w(k)] \qquad \text{(A-2-43a)}$$

$$v_{p\ell}(k) \triangleq [v_\ell(k-p),\ v_\ell(k-p+1),\ \cdots,\ v_\ell(k)] \qquad \text{(A-2-43b)}$$

$M_{p\ell}$ is the p-th order partial observability matrix given by

$$M_{p\ell} \triangleq \begin{bmatrix} C_\ell \\ C_\ell A_\ell \\ \vdots \\ C_\ell A_\ell^p \end{bmatrix} \quad \text{a } (p+1)m \text{ by } n \text{ matrix} \ . \qquad \text{(A-2-43c)}$$

$\Pi_{p\ell}$ is the extended process noise matrix of the form

$$\Pi_{p\ell} \triangleq \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ C_\ell D_\ell & 0 & 0 & \cdots & 0 & 0 \\ C_\ell A_\ell D_\ell & C_\ell D_\ell & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & & \vdots & \vdots \\ C_\ell A_\ell^{p-1} D_\ell & \cdot & \cdot & \cdots & C_\ell D_\ell & 0 \end{bmatrix} \quad \text{a } (p+1)m \text{ by } (p+1)n_w \text{ matrix} \quad . \tag{A-2-43d}$$

Note that the probability density of $Y_p(k)$ is a sum of L Gaussian densities:

$$p(Y_p(k)) = \sum_{\ell=1}^{L} P_\ell \, N(0, \hat{C}_\ell)$$

where $\hat{C}_\ell$ is the autocovariance function (ACF) of Y_p conditioned on the fact that the model hypothesis is ℓ. In the steady state, $\hat{C}_\ell$ is given by

$$\hat{C}_\ell = M_{p\ell} \Sigma_\ell M_\ell^T + \Pi_{p\ell} Q_p \Pi_{p\ell}^T + R_{p\ell} \tag{A-2-45}$$

where

$$R_{p\ell} \triangleq \text{diag } (R_\ell) \quad \text{a } (p+1)m \text{ by } (p+1)m \text{ matrix} \tag{A-2-46a}$$

$$Q_p \triangleq \text{diag } (Q) \quad \text{a } (p+1)n_w \text{ by } (p+1)n_w \text{ matrix} \tag{A-2-46b}$$

$$\Sigma_\ell = A_\ell \Sigma_\ell A_\ell^T + D_\ell Q D_\ell^T \quad . \tag{A-2-46c}$$

Using Eq. A-2-46c, repeatedly, it is easy to see that $\hat{C}_\ell$ is a (p+1) by (p+1) block Toeplitz matrix of the form:

$$\hat{C}_{\ell} = \begin{bmatrix} \hat{C}_{0\ell} & \hat{C}_{1\ell}^{T} & \cdots & C_{p\ell}^{T} \\ \hat{C}_{1\ell} & \hat{C}_{0\ell} & \cdots & \hat{C}_{p-1,\ell}^{T} \\ \vdots & \vdots & & \vdots \\ \hat{C}_{p\ell} & \hat{C}_{p-1,\ell} & \cdots & \hat{C}_{0\ell} \end{bmatrix} \qquad \text{(A-2-47a)}$$

$$\hat{C}_{i\ell} = \begin{cases} C_{\ell}\, \Sigma_{\ell}\, C_{\ell}^{T} + R_{\ell}; \; i=0 \\ \\ C_{\ell}\, A^{i}\, \Sigma_{\ell}\, C_{\ell}^{T} \; ; \; i=1,2,\cdots,p \end{cases} \qquad \text{(A-2-47b)}$$

where Σ_{ℓ} is the steady-state covariance matrix of the state x(k).

Our objective is to find the t best parity relations that are maximally insensitive to model uncertainties and the process and measurement noises. We discuss two robust redundancy metrics that accomplish this objective.

A-2.4.2 Covariance Based Robust Redundancy Metric

In [13] a "robust redundancy metric" is proposed that is a weighted average of the trace of the covariance of μ(k) under various model hypotheses, $\ell=1,2,\cdots,L$:

$$J(p,t) = \sum_{\ell=1}^{L} P_{\ell}\, \mathrm{tr}\, \left[E_{\ell} \left\{ \mu(k)\, \mu^{T}(k) \right\} \right] \qquad \text{(A-2-48)}$$

where E_{ℓ} denotes expectation conditioned on the ℓ-th model hypothesis (note also that x(k), w(k), v(k), and hence μ(k) are zero mean). In Eq. A-2-48, we have indicated the explicit dependence of J on the order of the parity relations, p, as well as the desired number of parity relations, t. Using

Eq. A-2-41a and performing some algebraic manipulations, we have the following minimization problem for the optimal choice of the (p+1)m by t parity transformation matrix G :

$$\min_{G} J(p,t) = \min_{G} \text{tr } (G^T \hat{C} G) \qquad \text{(A-2-49a)}$$

subject to

$$G^T G = I_t \quad . \qquad \text{(A-2-49b)}$$

The constraint in Eq. A-2-49b is included to avoid the trival solution of G=0. In Eq. A-2-49a, $\hat{C}$ is the average ACF of Y_p over the set of uncertain models, $\ell=1,2,\cdots,L$ and is given by

$$\hat{C} = \sum_{\ell=1}^{L} P_\ell \hat{C}_\ell \qquad \text{(A-2-50)}$$

where $\hat{C}_\ell$ is the ACF of Y_p conditioned on model ℓ, defined in Eqs. A-2-45 - A-2-47. The optimum transformation matrix G is given by the following theorem.

Theorem 2.3: The robust redundancy metric J(p,t) is minimized by taking the columns of G to be the t normalized eigenvectors associated with the smallest t eigenvalues of the average ACF $\hat{C}$ of Eq. A-2-50.

Proof: Form the augmented Lagrangian function:

$$J'(p,t) = \text{tr } [G^T \hat{C} G + (G^T G - I_t)\, \Gamma] \qquad \text{(A-2-51)}$$

where Γ is the t by t matrix of Langrange multipliers. The graident of J' with respect to G is given by

$$\nabla_G J' = 2\, (\hat{C} G + G\Gamma) \qquad \text{(A-2-52)}$$

At the optimum G^*, the gradient is zero. Using this condition and the constraint Eq. A-2-49b, we have

$$\Gamma = -\left(G^{*T}\ \hat{C}G^*\right) \quad . \tag{A-2-53}$$

Using Eq. A-2-53 in Eq. A-2-52, we have

$$\hat{C}G^* = G^*\left(G^{*T}\hat{C}G^*\right) \quad . \tag{A-2-54}$$

Equation A-2-54 implies that the optimum parity transformation matrix G^* is formed by taking the columns of G^* to be the t normalized eigenvectors associated with the smallest eigenvalues of the average ACF $\hat{C}$.

The results of Theorem 2.3 provide us with an important interpretation of parity relations. Since the parity vectors g_i are the eigenvectors of $\hat{C}$ associated with the minimum eigenvalues, the parity checks $\mu_i(k)$ of order p correspond to "near perfect" correlation among the output variables over a window of size p . In addition, since the eigenvalues are measures of signal energy along the directions represented by the eigenvectors, the parity vectors g_i can be thought of as the "directions of least output energy." Moreover, if the eigenvalues λ_i of $\hat{C}$ are ordered according to size

$$\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{(p+1)m} \tag{A-2-55}$$

then with the columns of G^* chosen as the eigenvectors corresponding to the smallest t eigenvalues, the optimum values of the robust redundancy metric is

$$J^*(p,t) = \sum_{i=1}^{t} \lambda_i \quad . \tag{A-2-56}$$

That is, the first column of G^*, g_1, is the most robust parity check and λ_1 is its measure of robustness, g_2 is the next best parity check with λ_2 as its

measure of robustness, and so on. Thus $J^*(p,t)$ has the interpretation as the overall robustness measure, if one were to use the t best parity relations of order p. Since the λ_i are increasing, the general shape of $J^*(p,t)$ as a function of p and t is as shown in Fig. A-2-2. What this curve provides us is a summary picture of the level of robustness in a particular set of parity relations. Intuitively, for a given p, a good choice of t would be near the "knee" of the curve (i.e., at values of t at which λ_{t+1} begins to increase more dramatically), although the value of t also depends on the number of failure modes that should be detected and isolated (see Section A-3).

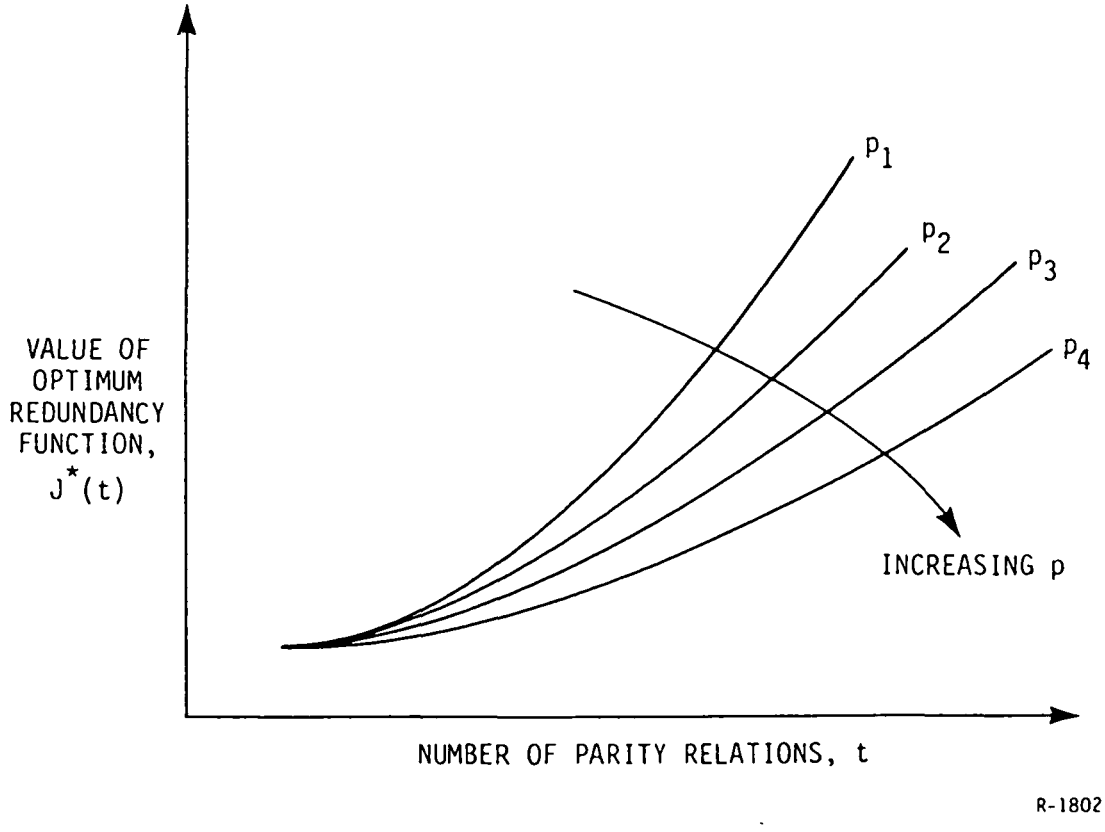


Figure A-2-2. Illustration of the Robust Redundancy Function

Another potential use of the robust redundancy curve is in the design process. Specifically, different sensor complements (i.e., different choices of the measurement matrix C_ℓ in Eq. A-2-38) will in general have different redundancy characteristics. By comparing their robust redundancy curves one can determine a useful measure for the relative merits of alternative sensor sets in terms of the likely FDI algorithm performance that would result in each case. Also it is worth noting that in the development in [13], the parity vectors are interpreted as the left singular vectors of a scaled and extended observability matrix rather than the eigenvector interpretation presented here. Our interpretation of parity check generation in terms of eigen-analysis of the average ACF function allows us to work with experimentally-obtained covariance data, if a model of the system is not available. In addition, as will be discussed in the following, our results provide an important link between parity check generation problem and the results in autoregressive (AR) spectral estimation [23]-[25], system identification [20], model order reduction [19] and the feature extraction problems in pattern recognition [22].

A-2.4.3 Entropy Based Redundancy Metric

An alternative metric for robust redundancy can be derived based on the concept of entropy. As discussed in Section 1, a parity check can be thought of as a signal that takes on values as close to a specified deterministic value (typically zero) as possible under normal system operation. The less randomness there is in the parity check, the more useful it is for detecting failures. Thus, a meaningful parity selection criterion is to choose parity relations $\mu_i(k)$ such that they have minimum entropy. We use a weighted average of the entropy under various model hypotheses $\ell=1,2,\cdots,L$ as our robust redundancy metric:

$$H_{\mu}(p,t) = \sum_{\ell=1}^{L} P_{\ell} H_{\mu|\ell} \qquad \text{(A-2-57)}$$

where $H_{\mu|\ell}$ is the entropy of $\mu(k)$ conditioned on model ℓ. Since $\mu(k)$ is Gaussian with zero mean and covariance $G^T\hat{C}_{\ell}G$ under the assumption of model ℓ, $H_{\mu|\ell}$ is given by

$$H_{\mu|\ell} = \frac{t}{2} \ell n\ (2\pi e) + \frac{1}{2} \ell n \det\ (G^T\hat{C}_{\ell}G) \qquad \text{(A-2-58)}$$

Thus, the optimization problem is to minimize $H_{\mu}(p,t)$ in Eq. A-2-57, with respect to the parity transformation G , subject to the normalization constraint $G^TG=I_t$. The result is given by Theorem 2.4.

Theorem 2.4: The entropy function $H_{\mu}(p,t)$ is minimized by a parity transformation matrix G^* that satisfies the nonlinear equation:

$$G^* = \sum_{\ell=1}^{L} P_{\ell} \hat{C}_{\ell} G^* (G^{*T}\hat{C}_{\ell}G^*)^{-1} \qquad \text{(A-2-59)}$$

Proof: Same as Theorem 2.1.

A successive substitution scheme can be used to solve for the optimal minimum entropy parity transformation. However, if we approximate the density function of $Y_p(k)$ (or $\mu(k)$), which is a sum of L Gaussian densities, by an equivalent single Gaussian density with the same mean and covariance, as in Eq. A-2-36, an important simplification arises. That is,

$$E\ \{\mu(K)\} = 0 \qquad \text{(A-2-60a)}$$

$$\text{cov}\ \{\mu(k)\} = G^T \left(\sum_{\ell=1}^{L} P_{\ell} \hat{C}_{\ell} \right) G = G^T\hat{C}G \qquad \text{(A-2-60b)}$$

The optimum parity transformation is given by the following theorem.

Theorem 2.5: The Gaussian sum approximation implies that an optimal choice of G is the set of t eigenvectors of the average ACF $\hat{C}$ corresponding to the t smallest eigenvalues.

Proof: Same as Theorem 2.1.

The result of Theorem 2.5 says that the implicit probabilistic assumption made in the work of Lou, Willsky and Verghese [13] is precisely the Gaussian sum approximation (Eq. A-2-60) and that their metric provides robust parity relations with respect to the "average" model defined by the approximation in Eq. A-2-60.

Example 2.3: The F-100 Engine Example

In order to determine the optimally redundant parity vectors under normal (unfailed) conditions, we used the linearized system matrices A and C from the simplified nonlinear simulation of Ref. [26] at power lever angles (PLA) of 60°, 65° and 70°. The control is assumed to be zero for simplicity. As will be illustrated subsequently, the effect of control on parity generation can be included into the parity generation technique in a straightforward manner. The three sets of system matrices A and C, shown in Table A-2-1, exhibit a significant change in parameters. In this example, state dimension, n=4 and output dimension m=5. The four state variables are:

$x_1 = N_1$ Fan speed (RPM)

$x_2 = N_2$ Compressor Speed (RPM)

x_3 = TT410 Burner exit slow response temperature (°K)

x_4 = TT4.510 Fan turbine inlet slow response temperature (°K)

TABLE A-2-1. SYSTEM MATRICES A AND C AS A FUNCTION OF POWER LEVEL ANGLES (PLA)

PLA-degrees	Matrix	Coefficient Values			
60°	A	-5.133 -0.298 -0.0034 .0284	4.750 -2.982 -0.00013 -0.0290	-0.323 1.244 -0.639 -0.0413	1.408 -3.341 .0033 -1.918
	C	1.000 0.000 0.150 0.302 -1.344	0.000 1.000 0.906 0.007 -0.447	0.000 0.000 -0.013 -0.023 -0.226	0.000 0.000 -0.0124 -0.0195 .0148
65°	A	-4,827 -0.0623 -0.0362 .0007	4.749 -2.887 -0.0013 -0.029	-0.310 1.195 -0.639 -0.0438	1.353 -3.131 .0026 -1.918
	C	1.000 0.000 0.195 .309 -1.328	0.000 1.000 0.878 .0077 -0.411	0.000 0.000 -0.0109 -0.022 -0.229	0.000 0.000 -0.011 -0.0183 -0.0122
70°		-4.549 -0.0966 -0.0386 -0.0229	4.749 -2.795 -0.0013 -0.029	-0.2969 1.147 -0.638 -0.046	1.299 -2.928 .002 -1.918
		1.000 0.000 .233 .3433 -1.309	0.000 1.000 .851 .008 -0.379	0.000 0.000 -0.0092 -0.021 -0.233	0.000 0.000 -0.0098 -0.0172 .0100

The five sensor outputs are:

$y_1 = N_1$ Fan speed (RPM)

$Y_2 = N_2$ Compressor speed (RPM)

Y_3 = PT4 = Burner pressure (N/m^2)

Y_4 = PT6 = Augmentor pressure (N/m^2)

Y_5 = FTIT = Fan turbine inlet temperature (^oK)

Our objective is to find a single set of 5 parity relations that are linear combinations of present and past sensor outputs, for the three different operating conditions, PLA = 60°, 65° and 75°. If we can successfully design such a set for these three widely differing operating conditions, it is anticipated that the parity vectors for smaller uncertainties (approximately ±3°PLA) around a single operating point will be substantially robust. Assuming $\Sigma_\ell = I$ and $R_\ell = 0$ for $\ell = 1,2,3$ operating conditions, the optimal robust redundancy function, $J^*(p,t)$ is computed for various values of t and $p = 0,1,2,3$ and plotted in Fig. A-2-3. It is clear that a complete set of five robust parity relations can be obtained by using parity orders of at most 2. The parity transformation matrix, G associated with the 5 parity relations is shown in Table A-2-2.

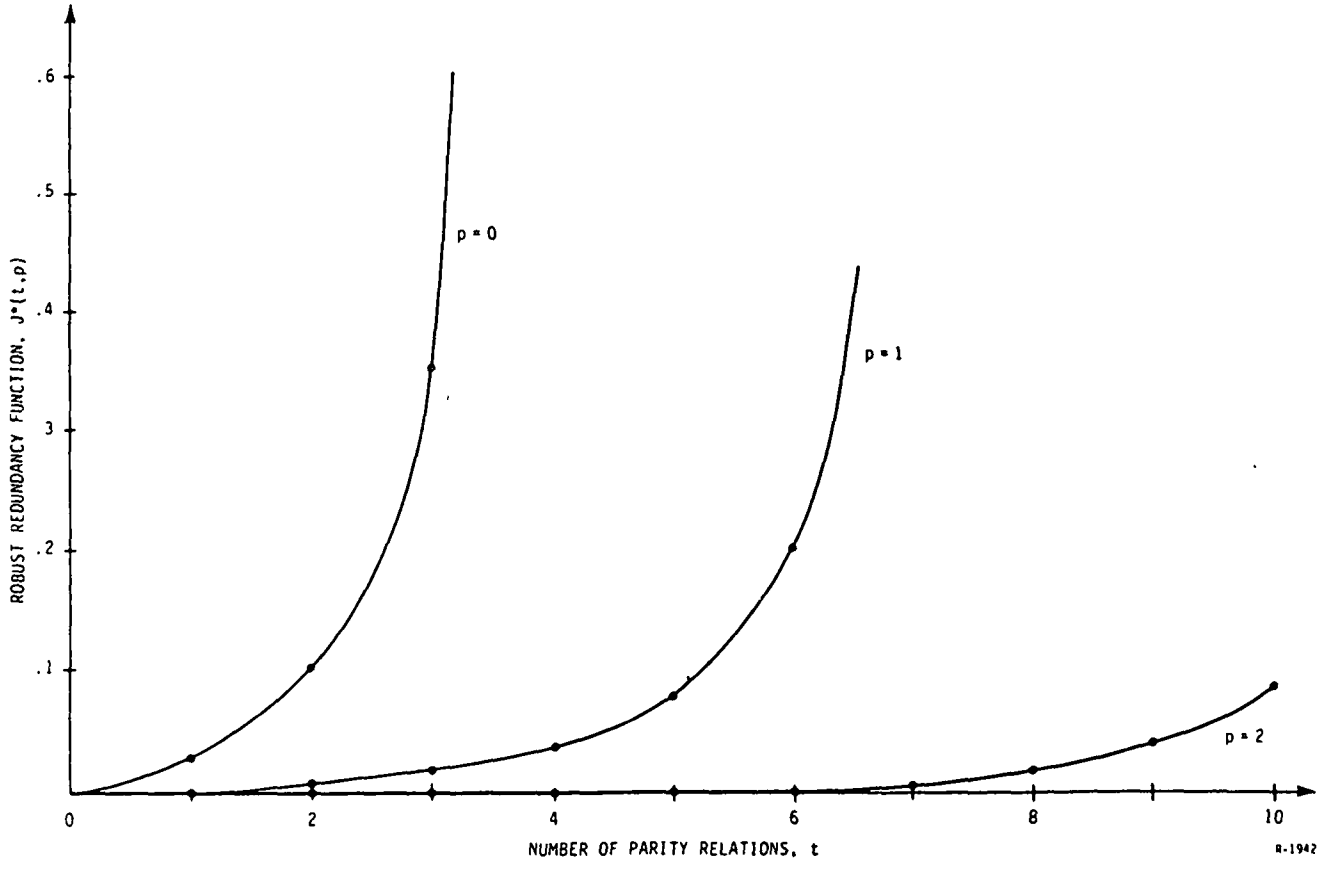


Figure A-2-3. Robust Redundancy Function vs. the Number of Parity Relations for the F-100 Example.

TABLE A-2-2. THE PARITY TRANSFORMATION MATRIX G FOR THE F-100 EXAMPLE

$$v = G \begin{bmatrix} Y(k-2) \\ Y(k-1) \\ Y(k) \end{bmatrix}$$

$$G = \begin{bmatrix}
.743 & .220 & -.045 & .0097 & .112 & .0564 & .179 & .009 & .0013 & -.571 & .100 & .059 & .0415 & .0705 & .0403 \\
-.182 & .0163 & -.0033 & .0054 & .0173 & .0723 & .221 & .0099 & .0016 & .0168 & .233 & .058 & -.0571 & .9041 & .185 \\
.057 & -.604 & .0923 & -.0267 & -.187 & .0490 & .1382 & .0045 & .0004 & -.093 & .0211 & .518 & -.151 & -.158 & .489 \\
.1636 & -.679 & .105 & -.033 & -.220 & .0094 & .0154 & -.0012 & -.0006 & -.137 & .0243 & -.487 & .153 & .1609 & -.380 \\
-.607 & .0606 & .0038 & .0034 & -.0071 & .092 & .279 & .0125 & .0020 & -.688 & .066 & -.131 & .0845 & -.194 & .0211
\end{bmatrix}$$

2.4.4 Complements

The parity check generation methodology discussed in subsections A-2.4.2 and A-2.4.3 has important implications in autoregressive spectral estimation [23-25], system identification and model order reduction [19,20]. Some of these issues, as well as frequency domain interpretations of parity relations, are discussed below.

Parity Relations as Reduced Order ARMA Models

Recall from Eq. A-2-39 that the component i of $\mu(k)$ can be written as:

$$\mu_i(k) = \sum_{j=0}^{p} g_{ij}^T y(k-p+j) = g_i^T Y_p(k) \quad ; \; i=1,2,\cdots,t$$

where g_i is column i of G . We say that the parity relation is strictly of order p if $g_{ip}^T \neq 0$. Consider such a relation and denote the components of g_{ip} by

$$g_{ip} = [g_{ip1}, g_{ip2}, \cdots, g_{ipq}, \cdots, g_{ipm}] \; . \qquad \text{(A-2-61)}$$

Assuming, without loss of generality, that $g_{ipq} \neq 0$, then component q of output vector, $y_q(k)$, can be predicted from $y_q(k-1)$, $y_q(k-2)$, $\cdots$, $y_q(k-p)$ and other output components $y_r(k)$, $y_r(k-1)$, $\cdots$, $y_r(k-p)$, $r \neq q$ using the i-th parity relation as:

$$\hat{y}_q^{(i)}(k) = \frac{-1}{g_{ipq}} \left[\sum_{j=0}^{p-1} g_{ijq} \, y_q(k-p+j) + \sum_{j=0}^{p} \sum_{\substack{r=1 \\ r\neq q}}^{m} g_{ijr} \, y_r(k-p+j) \right] \qquad \text{(A-2-62)}$$

where the superscript i denotes that y_q is predicted using the i-th parity relation. There are two important interpretations of Eq. A-2-62. The most obvious is that the right-hand side of Eq. A-2-62 represents a synthetic

measurement of the q-th output that can be compared to $y_q(k)$ to generate (a scaled version of) the parity relation $\mu_i(k)$. Thus Eq. A-2-62 can be interpreted as a prediction filter and leads to a generalization of the concept of parameter synthesis, discussed in [16] and successfully applied in [4], where the generalization stems from allowing the right-hand side of the prediction system in Eq. A-2-62 to have memory. The second interpretation of Eq. A-2-62 is as a reduced order ARMA model for $y_q(k)$. That is, $y_q(k)$ is expressed in terms of its own lagged values and of the present and past values of the remaining output components y_r, r=1,2,•••,m, r≠q, which are treated as external inputs, and a noise process $\nu_i(k)$, i.e.,

$$y_q(k) = \frac{-1}{g_{ipq}} \left[\sum_{j=0}^{p-1} g_{ijq}\, y_q(k-p+j) + \sum_{j=0}^{p} \sum_{\substack{r=1 \\ r\neq q}}^{m} g_{ijr}\, y_r(k-p+j) - \mu_i(k) \right] . \qquad \text{(A-2-63)}$$

This method of obtaining reduced order ARMA models has similarities to the work in [19], [20].

Parity Relations as Prediction Error Filters

From Eqs. A-2-62 and A-2-63, it is clear that the parity relation $\mu_i(k)$ can be regarded as a (weighted) one-step prediction error filter (PEF) of order p, since

$$\mu_i(k) = g_{ipq} \left[y_q(k) - \hat{y}_q^{(i)}(k) \right] . \qquad \text{(A-2-64)}$$

Since the PEF has finite memory, $\mu_i(k)$ is also referred to as a finite memory (FM) parity check [12]. PEF interpretation of parity relations is shown in Fig. A-2-4. PEFs have been studied extensively in the AR spectral estimation

literature [23]-[25]. Our method of obtaining PEF coefficients is akin to Pisarenko's harmonic retrieval method discussed in [23]-[25]. However, unlike the approches of spectral estimation literature, the parity space approach does not fix the dimension of $\mu(k)$ and, hence, the dimension of the PEFs a priori. The number of parity relations, t is chosen to minimize the cost function (Eq. A-2-48 or Eq. A-2-57) while satisfying the FDI requirements (see Section 3). In addition, in computing the prediction error for $y_q(k)$, all other output components $y_r(k)$, $r \neq q$ are treated as known inputs to the ARMA model. Finally, model uncertainty is taken into account explicitly in computing the average ACF that forms the basis for designing robust parity coefficients, whereas the covariance based PEFs treated in the spectral estimation literature assume a perfectly known model.

Parity Relations as Moving Average (MA) Filters

The window of observations $Y_p(k)$ can be written in the frequency domain as

$$Y_p(z) = \Delta_m(z)\, y(z) \qquad \text{(A-2-65)}$$

where $y(z)$ is the z transform of the output, $y(k)$; and $\Delta_m(z)$ is a $(p+1)m$ by m matrix of delay elements given by

$$\Delta_m^T(z) \triangleq [z^{-p}I_m, z^{-(p-1)}I_m, \cdots, I_m] \ . \qquad \text{(A-2-66)}$$

Then, the parity check vector $\mu(k)$ can be written in the frequency domain as:

$$\mu(z) = G^T Y_p(z) = G^T(z)\, y(z) \qquad \text{(A-2-67)}$$

where

$$G^T(z) = G^T \Delta_m(z) \qquad \text{(A-2-68)}$$

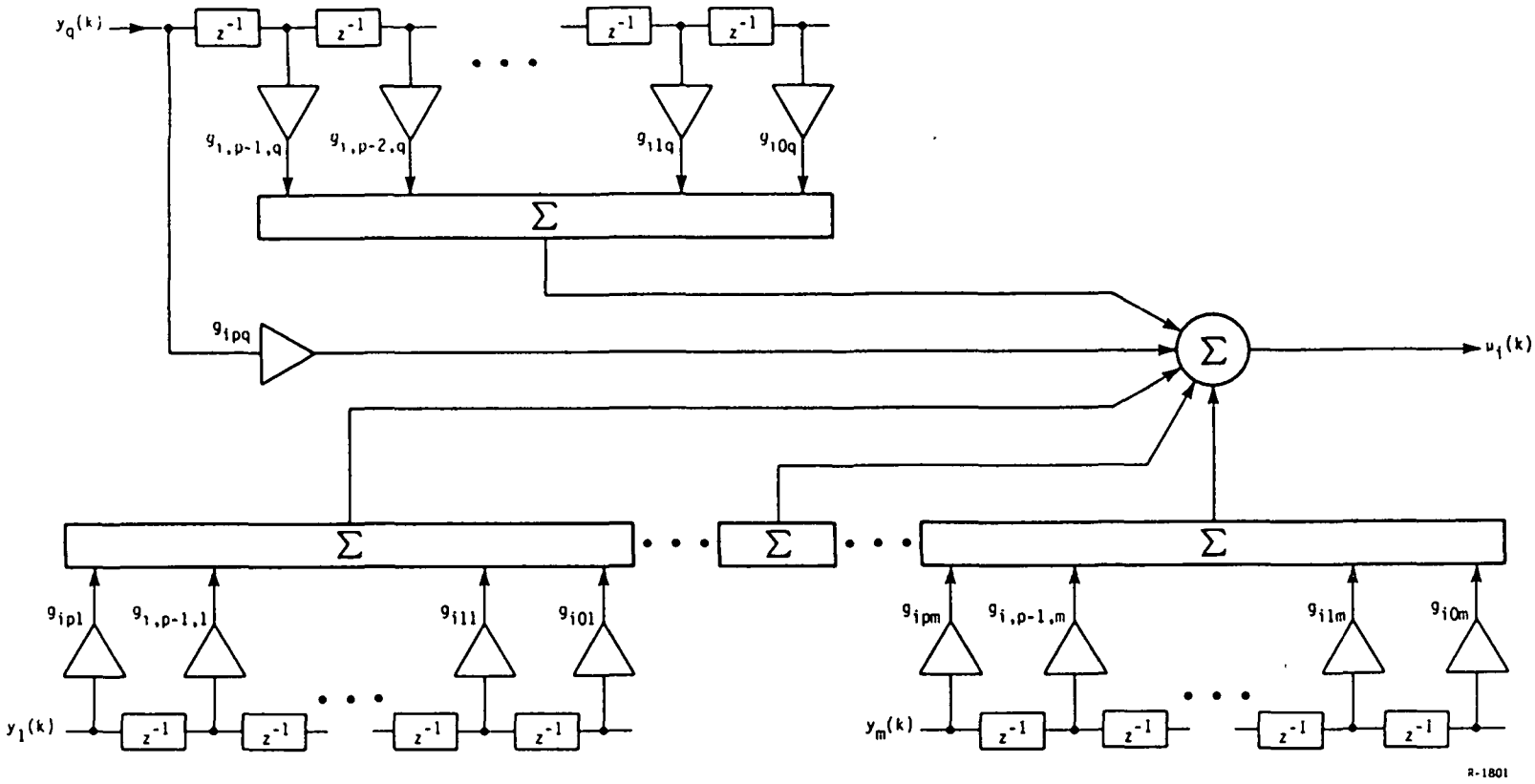


Figure A-2-4. Prediction Error Filter Interpretation of Parity Relations

If we partition the (p+1)m by t parity transformation matrix G as

$$G^T = [G_0^T \; G_1^T, \ldots, G_p^T] \qquad \text{(A-2-69)}$$

where each G_j is an m by t matrix given by

$$G_j = [g_{1j} \; g_{2j}, \ldots, g_{tj}] \; ; \quad 0 \leq j \leq p \quad . \qquad \text{(A-2-70)}$$

The m dimensional column vectors g_{ij}, $1 \leq i \leq t$ are defined in Eq. A-2-61. Using the definition of $\Delta_m(z)$ in Eq. A-2-66, we have

$$\mu(z) = \sum_{j=0}^{p} G_j^T \, z^{-(p-j)} y(z) = G^T(z) \; y(z) \quad . \qquad \text{(A-2-71)}$$

Thus, the parity transformation is a set of t moving average (MA) filters of order p (in m dimensional space), as shown in Fig. A-2-5.

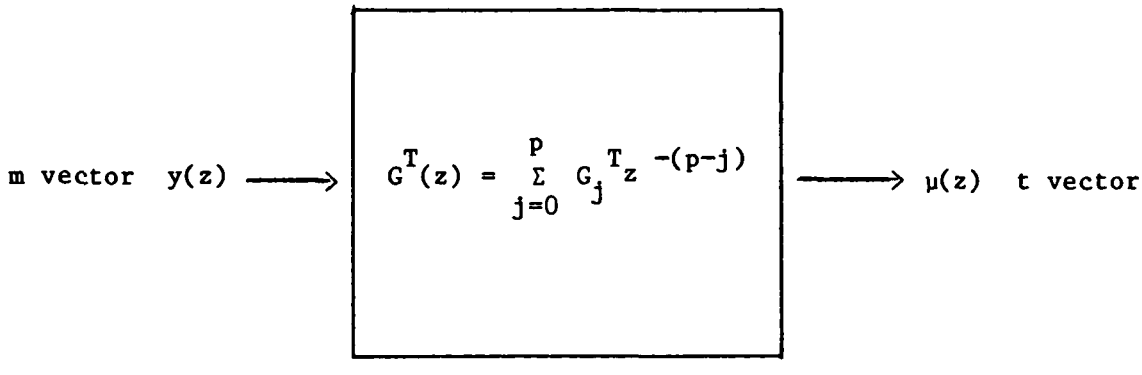


Figure A-2-5. Parity Transformation as a Moving Average (MA) Filter

A Frequency Domain Algorithm for Parity Transformation

The MA filter interpretation provides us with a method for computing the parity transformation G in the frequency domain. In the steady state letting $z = e^{j\omega}$, we have

$$\mu(e^{j\omega}) = G^T(e^{j\omega})\, y(e^{j\omega}) \tag{A-2-72}$$

where $G(e^{j\omega})$ is an m by t impulse response matrix. If we let $\Phi_\mu(\omega)$ be the average power spectral density (PSD) matrix (which is a real function of ω) over all possible model hypotheses, then the autocovariance related criterion in Eq. A-2-48 becomes:

$$J(p,t) = \text{tr}\left[(2\pi)^{-1} \int_0^{2\pi} \Phi_\mu(\omega)\, d\omega\right] \tag{A-2-73}$$

Equation A-2-73 is derived by using the property that the average autocovariance function (ACF) of $\mu(k)$, $G\,\tilde{C}G^T$ and the PSD $\Phi_\mu(\omega)$ are Fourier transform pairs. The average PSD matrix $\Phi_\mu(\omega)$ is given by

$$\Phi_\mu(\omega) = \sum_{\ell=1}^{L} p_\ell\, \Phi_{\mu\ell}(\omega) \tag{A-2-74}$$

where $\Phi_{\mu\ell}(\omega)$ is the PSD of $\mu(k)$ under model ℓ and is related to the PSD of the output process $\Phi_{y\ell}(\omega)$ via

$$\Phi_{\mu\ell}(\omega) = G^T(e^{j\omega})\, \Phi_{y\ell}(\omega)\, G(\bar{e}^{j\omega}) \tag{A-2-75}$$

The PSD of the output process is related to the system parameters under model ℓ and is given by

$$\Phi_{y\ell}(\omega) = C_\ell(e^{j\omega I} - A_\ell)^{-1} D_\ell\, Q\, D_\ell^T\, (e^{-j\omega I} - A_\ell)^{-1} + R_\ell \tag{A-2-76}$$

Equation A-2-73 must be minimized subject to the constraint:

$$\int_0^{2\pi} G^T(e^{j\omega})\, G(\bar{e}^{j\omega})\, d\omega = 2\pi\, I_t \qquad \text{(A-2-77)}$$

It is easy to see that the best choice for $G(e^{j\omega})$, (although not necessarily causal), at frequency ω is the set of t eigenvectors corresponding to the smallest t eigenvalues of the average PSD $\Phi_y(\omega)$ given by

$$\Phi_y(\omega) = \sum_{\ell=1}^{L} p_\ell\, \Phi_{y\ell}(\omega) \qquad \text{(A-2-78)}$$

An approximate algorithm for obtaining a pth order MA filter is as follows:

a. Divide the interval $[0,2\pi]$ into p equal segments, and let

$$\omega_j = j \cdot \frac{2\pi}{p} \quad ; \quad j = 0,1,2,\cdots,p$$

b. Evaluate $\Phi_{y\ell}(\omega)$ for each $\ell = 1,2,\ldots,L$ and the t orthogonal eigenvectors, G_j corresponding to the smallest t eigenvalues of the average PSD $\Phi_y(\omega_j)$. Denote them by $\lambda_i(\omega_j)$, $i=1,2,\cdots,t$. Scale each eigenvector by $\lambda_i^{-0.5}(\omega_j)$.

c. Normalize each column of G so that

$$\sum_{j=0}^{p} g_{ij}^T g_{ij} = 1 \quad ; \; i=1,2,\cdots,p$$

where g_{ij} is the scaled version of the i-th column of G_j (an m vector). The resulting normalized G_j provide the desired MA filter coefficients. The procedure is only approximate since G_js are computed from only a finite set of ω's. However, it has the advantage that the eigenvalue - eigenvector decomposition of an m by m matrix, G_j, need be performed (p+1) times rather than the eigenvalue - eigenvector decomposition of a (p+1)m by (p+1)m matrix as in Eq. A-2-49. Since the approximation method is computationally efficient, it

is not necessary to restrict the value of p to small numbers. An alternate method is to find G_j, $j = 0,1,2,\cdots,N$ ($N \gg 1$) using the eigenanalysis and then perform inverse FFT to find the impulse response matrix G. One drawback of this approach is that the resulting G need not be nonanticipatory [20].

Another major advantage of the MA filter interpretation of parity relations is that it provides us with a general method for generating parity vectors by weighting frequency ranges of interest. Suppose, W(z) represents such an m by m weighting filter so that the transformed output $\tilde{y}(z)$ is given by

$$\tilde{y}(z) = W(z)\ y(z) \tag{A-2-79}$$

For a given W(z), we want to find a pth order parity transformation (i.e., an MA filter) so that $\mu(z) = G^T(z)\ \tilde{y}(z)$ has minimum autocovariance or minimum entropy. It is easy to see that the eigen analysis of the average PSD of $\tilde{y}(k)$, $\tilde{\Phi}_y(\omega)$ at $\omega_j = j \cdot \frac{2\pi}{p}$ for $j = 0,1,2,\cdots,p$ provides us with (approximate) p-th order MA coefficients, G_j, where $\tilde{\Phi}_y(\omega)$ is given by

$$\tilde{\Phi}_y(\omega) = W(e^{j\omega})\ \Phi_y(\omega)\ W(\bar{e}^{j\omega}) \tag{A-2-80}$$

Generalization of Levinson Filter

If, instead of $G^TG = I_t$, we impose a constraint on the coefficients multiplying y(k) as

$$G_p{}^T G_p = I_t \tag{A-2-81}$$

$$G_p{}^T = [g_{1p}, g_{2p}, \cdots, g_{tp}] \tag{A-2-82}$$

is an mxt matrix, then we obtain a generalization of Levinson filter. Note that the Levinson filter imposes the constraint $G_p = I_m$. The constraint in

Eq. A-2-81 says that we require the best t relations of the m vector output that minimizes the robustness metric in Eq. A-2-49. The optimal G is given by the following theorem.

Theorem 2.6

The robustness metric (Eq. A-2-49) is minimized subject to the constraint in Eq. A-2-81 by choosing the columns of G_p^* as the set of t eigenvectors of the matrix $\hat{C}_0$ given by

$$\hat{C}_0 = \tilde{C}_0 - N_p{}^T O_{p-1}^{-1} N_p \qquad \text{(A-2-83a)}$$

where

$$N_p^T \triangleq [\, \hat{C}_{p-1}, \hat{C}_{p-2}, \cdots, \hat{C}_1 \,] \quad \text{an m by pm matrix} \qquad \text{(A-2-83b)}$$

$$O_{p-1} \triangleq \begin{bmatrix} \hat{C}_0 & \hat{C}_1{}^T & \cdots & \hat{C}_{p-1}{}^T \\ \hat{C}_1 & \hat{C}_0 & \cdots & \hat{C}_{p-2}{}^T \\ \vdots & \vdots & & \\ \hat{C}_{p-1} & \hat{C}_{p-2} & \cdots & \hat{C}_0 \end{bmatrix} \quad \text{an mp by mp matrix} \qquad \text{(A-2-83c)}$$

where $\hat{C}_0, \hat{C}_1, \cdots, \hat{C}_p$ are the average ACFs defined in Eqs. A-2-47 and A-2-50. That is,

$$\hat{C}_i = \sum_{\ell=1}^{L} P_\ell \hat{C}_{i\ell} \quad ; \quad 0 \leq i \leq p \quad . \qquad \text{(A-2-83d)}$$

The remaining G_j are given by

$$[G_0^{*T} G_1^{*T}, \cdots, G_{p-1}^{*T}] = -G_p^{*T} N_p^T O_{p-1}^{-1} \qquad \text{(A-2-83e)}$$

Proof: The augmented Lagrangian function is

$$J'(p,t) = \mathrm{tr}\ [G^T \hat{C} G + (G^T E E^T G - I_t)\ \Gamma] \qquad \text{(A-2-84)}$$

where Γ is a t by t matrix of Lagrange multipliers and

$$E = \begin{bmatrix} 0 & 0 \cdots 0 \\ 0 & 0 \cdots 0 \\ \vdots & \\ 0 & 0 \cdots I_m \end{bmatrix} \quad \text{a } (p+1)m \text{ by } (p+1)m \text{ matrix}$$

The optimality condition $\nabla_G J' = 0$ at the optimal G^* gives

$$\hat{C}\ G^* + E\ E^T\ G^*\ \Gamma^* = 0 \qquad \text{(A-2-85a)}$$

or in expanded form

$$\begin{bmatrix} O_{p-1} & N_p \\ N_p^T & \hat{C}_0 \end{bmatrix} \begin{bmatrix} \hat{G}^*_{p-1} \\ G^*_p \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ -G^*_p \Gamma^* \end{bmatrix} \qquad \text{(A-2-85b)}$$

where

$$\hat{G}^{*T}_{p-1} \triangleq [G^{*T}_0\ G^{*T}_1 \cdots G^{*T}_{p-1}] \ . \qquad \text{(A-2-86)}$$

Equation A-2-85b implies that

$$\hat{G}^*_{p-1} = -\ O^{-1}_{p-1}\ N_p\ G^*_p \qquad \text{(A-2-87)}$$

and

$$N_p^T\ \hat{G}_{p-1} + \hat{C}_0\ G^*_p = -\ G^*_p\ \Gamma^* \ . \qquad \text{(A-2-88)}$$

Using the constraint $G_p^T G_p = I_t$, Γ^* is evaluated as

$$\Gamma^* = - G_p^{*T} \hat{C}_0 G_p^* \qquad \text{(A-2-89)}$$

where $\hat{C}_0$ is defined in Eq. A-2-83a. Equations A-2-88 and A-2-89 provide the desired result that the optimal G_p is formed from the set of t eigenvectors of $\hat{C}_0$ corresponding to its smallest t eigenvalues. Once G_p is computed, G_{p-1} is evaluated from Eq. A-2-86. It is worth noting that the matrix $\hat{C}_0$ arises prominently in the inversion of block Toeplitz matrices [27].

A Recursive Expression for the Time Window of Measurements

We can derive a recursive expression for $Y_p(k)$ from the system equations (Eqs. A-2-37 and A-2-42) whenever the system is observable. When $M_{p\ell}$ has full column rank, i.e., rank $[M_{p\ell}] = n$ under model hypothesis ℓ, then Eq. A-2-42 can be rewritten as:

$$x(k) = T_{p\ell} \left[Y_p(k) - \Pi_{p\ell} w_p(k) - v_{p\ell}(k) \right] \qquad \text{(A-2-90)}$$

where $T_{p\ell}$ is the generalized inverse of $M_{p\ell}$ given by

$$T_{p\ell} = \left(M_{p\ell}^T M_{p\ell} \right)^{-1} M_{p\ell}^T \ . \qquad \text{(A-2-91)}$$

Using Eq. A-2-90 in Eq. A-2-37, we have

$$\left[T_{p\ell} Y_p(k+1) - A_\ell T_{p\ell} Y_p(k) \right] = \left[T_{p\ell} \Pi_{p\ell} w_p(k+1) - A_\ell T_{p\ell} \Pi_{p\ell} w_p(k) \right]$$

$$+ \left[T_{p\ell} v_{p\ell}(k+1) - A_\ell T_{p\ell} v_{p\ell}(k) \right] + D_\ell w(k) \ . \qquad \text{(A-2-92)}$$

Since Eq. A-2-92 is in the form of a linear least-squares estimation problem for noise processes, Ref. [28] uses it as a basis to identify the noise

statistics (i.e., mean and variance of w and v). It may be noted that a direct recursive expression for $Y_p(k)$ can be obtained from Eqs. A-2-37, A-2-42 and A-2-90 as:

$$Y_p(k+1) = M_{p\ell}\, A_\ell\, T_{p\ell}\, Y_p(k) + M_{p\ell}\, D_{\ell w}(k)$$
$$+ \left[\Pi_{p\ell}\, w_p(k+1) - M_{p\ell}\, A_\ell\, T_{p\ell}\, \Pi_{p\ell}\, w_p(k)\right]$$
$$+ \left[v_{p\ell}(k+1) - M_{p\ell}\, A_\ell\, T_{p\ell}\, v_{p\ell}(k)\right] \quad . \qquad \text{(A-2-93)}$$

The advantage of Eq. A-2-92 over Eq. A-2-93 is that the estimation equation is of dimension n only (rather than (p+1)m of Eq. A-2-93).

Extension to Include Control Signals

We can easily extend the parity generation approach to systems involving control variables u(k). Formally, assume that the system dynamics are given by

$$x(k+1) = A_\ell\, x(k) + B_\ell\, u(k) + D_\ell\, w(k) \qquad \text{(A-2-94)}$$

$$y(k) = C_\ell\, x(k) + F_\ell\, u(k) + v_\ell(k) \qquad \text{(A-2-95)}$$

where x(k), $w_\ell(k)$, $v_\ell(k)$ and y(k) are as defined in Eqs. A-2-37 and A-2-38. u(k) is an n_u vector of control variables. As before, in order to generate a pth order parity relation, we form a window of observation $Y_p(k)$, satisfying

$$Y_p(k) = M_{p\ell}\, x(k) + \Gamma_{p\ell}\, U_p(k) + \Pi_{p\ell}\, w_{pi}(k) + v_\ell(k) \qquad \text{(A-2-96)}$$

where $Y_p(k)$, $M_{p\ell}$, $\Pi_{p\ell}$, $w_p(k)$ and $v_{p\ell}(k)$ are as defined in Eqs. A-2-40 and A-2-42. The control related matrix, $\Gamma_{p\ell}$ and extended control vector $U_p(k)$ are given by

$$\Gamma_{p\ell} \triangleq \begin{bmatrix} F_\ell & 0 & \cdots & 0 \\ C_\ell B_\ell & F_\ell & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ C_\ell A_\ell^{p-1} B_\ell & C_\ell A_\ell^{p-2} B_\ell & \cdots & F_\ell \end{bmatrix} \qquad \text{(A-2-97a)}$$

$$U_p^T(k) \triangleq [u(k-p)\ u(k-p+1)\ \cdots\ u(k)] \ . \qquad \text{(A-2-97b)}$$

Depending on whether the control signals are accessible for measurement or not, we can formulate two parity generation problems.

Problem 1: (Control Signals are not Accessible for Measurement)

Find a (p+1)m by t orthonormal parity transformation matrix G such that $tr[G^T \hat{C} G]$ is minimized, where

$$\hat{C} = \sum_{\ell=1}^{L} P_\ell E_\ell \{Y_p(k)\ Y_p^T(k)\} \ . \qquad \text{(A-2-98)}$$

The average ACF $\hat{C}$ can easily be evaluated, once the ACF of U_p denoted by $\hat{C}_{u\ell}$ under each of the model hypotheses ℓ, $\ell=1,2,\ldots,L$ is known. The solution of this problem involves finding the t eigenvectors of $\hat{C}$ corresponding to the t smallest eigenvalues.

Problem 2: (Control Signals are Accessible for Measurement)

Find a $(p+1)(m+n_u)$ by t orthonormal parity transformation matrix G such that $tr[G^T \hat{C}_a G]$ is minimized, where $\hat{C}_a$ is the average ACF of the augmented vector of outputs and control variables $Y_{pa}(k)$ given by

$$Y_{pa}^T(k) \triangleq [Y_p(k)\ U_p(k)] \qquad \text{(A-2-99a)}$$

$$\hat{C}_a(k) \triangleq \sum_{\ell=1}^{L} P_\ell \, E_\ell \{Y_{pa}(k) \, Y_{pa}^T(k)\} \quad . \tag{A-2-99b}$$

The average augmented ACF $\hat{C}_a$ is of the form:

$$\hat{C}_a = \begin{bmatrix} \hat{C} & \hat{C}_{yu} \\ \hat{C}_{yu}^T & \hat{C}_u \end{bmatrix} \tag{A-2-99c}$$

where

$$\hat{C}_{yu} = \sum_{\ell=1}^{L} P_\ell \, \Gamma_{p\ell} \, \hat{C}_{u\ell} \tag{A-2-99d}$$

$$\hat{C}_u = \sum_{\ell=1}^{L} P_\ell \, \hat{C}_{u\ell} \quad . \tag{A-2-99e}$$

The solution of problem 2 involves finding the t eigenvectors of $\hat{C}_a$ corresponding to the t smallest eigenvalues.

It may be noted that the key to the solution of problems 1 and 2 lies in the computation of the ACF of $U_p(k)$ under each model hypothesis ℓ, $\hat{C}_{u\ell}$. The ACF of $U_p(k)$ can, for example, be obtained in one of the following three ways:

a. u(k) is given by the feedback control law:

$$u(k) = k_p \, x(k) \quad . \tag{A-2-100}$$

b. u(k) is given by the proportional-integral feedback control law:

$$u(k) = k_1 \, y(k) + k_2 \sum_{i=0}^{k-1} y(i) \quad . \tag{A-2-101}$$

c. u(k) is an unknown, but bounded function or equivalently, $U_p(k)$ has known autocovariance function, $\hat{C}_{u\ell}$ under each model hypothesis ℓ.

A-2.5 SUMMARY

In this section, minimum entropy concept was used to provide a unified view of parity generation encompassing static and dynamic systems with and without noise and model uncertainty. Using this concept, an explicit non-linear equation (Eqs. A-2-34 and A-2-59) is derived for the robust parity transformation matrix, G in the presence of model uncertainties. In order to simplify computations, we have approximated the sum of Gaussian densities with a single Gaussian density with the same mean and covariance. This assumption led to a simple interpretation of G in terms of the eigenvectors corresponding to the smallest t eigenvalues (t = required number of parity relations) of the weighted sum of covariance matrices under various model hypotheses. This result was contrasted with the previous results of Chow, Lou, Willsky and Verghese [12],[13]. In addition, we derived a frequency domain algorithm to compute the parity transformation and have provided numerous interpretations of parity relations. Some of these interpretations are as follows:

1. For a deterministic system with no model uncertainty, the parity relations lie in the unobservable subspace under normal system conditions. That is, the range space of parity vectors is orthogonal to the range space of the observability matrix.

2. The parity relations can be viewed as reduced order ARMA models of individual outputs.

3. Parity relations can be interpreted as the prediction error filters (PEFs) that arise in the spectral estimation literature.

4. Parity relations correspond to near perfect correlations among measured outputs.

5. Parity relations are the directions of least energy.

6. Parity relations have minimum entropy (or the greatest certainty or robustness).

A-3. REDUNDANCY RELATIONS FOR ROBUST FAILURE DETECTION AND ISOLATION

A-3.1 PROBLEM DESCRIPTION AND MOTIVATION

The methods of Section A-2 do not take any specific failure mode into account but are aimed at selecting a set of robust parity relations that provide the smallest values for certain metrics under normal operating conditions and model uncertainty. However, the presence of robust redundancy in the sense of the previous discussion does not guarantee that all the failure modes can be detected and distinguished. In addition, the information about the presence or absence of a particular failure mode generally accumulates over time. Failure decisions making use of a time sequence of parity relations and/or measured system outputs will have lower error probabilities than "single-shot" decisions, if the proper use of failure information is made. In this section, we develop statistical distance-based criteria for optimally robust parity generation for the detection and isolation of specific failures and for the information collection phase of the FDI system.

The statistical distance based criteria serve four major roles in our robust FDI formalism. First, if the failure modes are strongly observable (in the sense of achieving a prescribed probability of error for a specified detection delay) through the use of the robust parity relations generated in Section A-2, then the robust FDI criteria of this section specify the optimal method of combining the parity relations over the detection window. That is, information collection can be conceptualized as the generation of a single

detection parity check (or "detection decision statistic") from the robust parity relations of Section A-2. Second, if certain failure modes are not strongly observable through the robust parity relations, then the robust FDI criteria of this section can be used to generate detection parity checks ("detection decision statistics") from the system output. The parity checks so generated are not optimally robust in the sense discussed in Section A-2, but are sensitive to a particular failure mode. Third, if one or more failure modes are not distinguishable through the robust parity relations and detection parity checks, the robust FDI measures can be used to generate isolation parity checks ("isolation decision statistics") that are most sensitive to distinguish the particular pair or pairs of failure modes. In this case, the operation of an isolation parity check is triggered by the corresponding detection parity checks. Finally, the statistical distance measure introduced in this section provide bounds on the probability of error, and, hence, can be used to evaluate the effectiveness of alternative FDI schemes in terms of their ability to detect and isolate various failure modes. Thus, the measures of this section provide a unified methodology to devise a "divide and conquer" approach to the robust FDI design problem. In this section, we restrict our attention to the second and fourth roles, viz., the generation of detection parity checks from the window of system measurements, Y_p, and the evaluation of alternative FDI schemes. However, the theory is applicable for the other two applications mutatis mutandis.

The specific problem we consider here is the choice of parity checks for the robust detection of a particular failure mode. We assume that the system model under normal operation is

$$x(k+1) = A_{N\ell}\, x(k) + D_{N\ell}\, w(k) \qquad \text{(A-3-1)}$$

$$y(k) = C_{N\ell}\, x(k) + v(k) \qquad \text{(A-3-2)}$$

while the system model under failure is

$$x(k+1) = A_{F\ell}\, x(k) + D_{F\ell}\, w(k) + d_F(k) \qquad \text{(A-3-3)}$$

$$y(k) = C_{F\ell}\, x(k) + v_F(k) + b_F(k) \qquad \text{(A-3-4)}$$

where $d_F(k)$ and $b_F(k)$ account for additive failure effects (e.g., biases, drifts). In this case, we would like to find parity checks $\mu(k)$ that result in maximally dispersed statistics under failed and normal modes of system operation.

Ideally, the optimum parity relations are those that minimize the probability of making erroneous decisions. However, in most situations an analytic expression for the probability of error is difficult to derive, and, even if it can be found, the expression is too complicated for optimization. Therefore, it is useful to search for criteria that are easier to evaluate and optimize and that are, in some sense, related to the probability of error. Statistical distance measures such as the I-divergence, the J-divergence, and the Bhattacharyya distance between two probability distributions under two hypotheses (normal and a failure mode or two different failure modes) provide such easily computable criteria. These distance measures have found wide applicability in several areas of systems engineering, and most notably in pattern recognition [17],[22], control systems [29]-[31], communications [32] and information theory [33],[34]. They also have a long history and utility in statistics [35]-[39]. In this section we propose criteria based on

J-divergence and the Bhattacharyya distance to generate parity relations to achieve maximum discrimination among failure modes. Before, we solve the parity generation problem, we provide a brief background on the distance measures and some of their useful properties.

A-3.2 THE DIVERGENCE AND THE BHATTACHARYYA DISTANCE MEASURES

J-divergence is a measure of dissimilarity between two hypotheses or probability distributions. To introduce the concept of J-divergence, consider two hypotheses N (normal) and F (failed), and let μ be a t vector of parity relations. We denote by $p(\mu/N)$ and $p(\mu/F)$ be the two conditional probability density functions of μ under normal and failed hypotheses respectively. Then, the discriminating information for hypothesis F over hypothesis N is given by the log likelihood ratio, L_{FN}:

$$L_{FN} = \ell n \frac{p(\mu/F)}{p(\mu/N)} \quad . \tag{A-3-5}$$

The average discriminating information for hypothesis F is defined by

$$I_{FN}(p,t) = \int_{\mu} p(\mu|F) \ \ell n \frac{p(\mu|F)}{p(\mu|N)} \ d\mu \tag{A-3-6}$$

where we have shown the explicit dependence of the average discriminating information on the order and the number of parity relations. The distance measure I_{FN} is often called the Kullback-Leibler (K-L) information measure or the I-divergence. The discriminating information for hypothesis N versus hypothesis F is given by the log likelihood ratio:

$$L_{NF} = \ell n \frac{p(\mu/N)}{p(\mu/F)} = - L_{FN} \tag{A-3-7}$$

The average discriminating information for hypothesis N is given by the K-L measure:

$$I_{NF(p,t)} = \int_{\mu} p(\mu/N) \ \ell n \frac{p(\mu/N)}{p(\mu/F)} \ d\mu \ . \qquad \text{(A-3-8)}$$

Note that $I_{FF} = I_{NN} = 0$. However, in general, the I-divergence is not symmetric, i.e.,

$$I_{FN} \neq I_{NF} \ . \qquad \text{(A-3-9)}$$

The total average information for distinguishing hypothesis F from hypothesis N is the J-divergence, first introduced into the statistical literature by Jeffreys [40,41]. The J-divergence is defined as the difference in the average values of the log likelihood ratio:

$$J_{FN(p,t)} = \int_{\mu} \left[p(\mu/F) - p(\mu/N)\right] \ \ell n \frac{p(\mu/F)}{p(\mu/N)} \ d\mu \ . \qquad \text{(A-3-10)}$$

It is easy to see from Eqs. A-3-6 and A-3-10 that

$$J_{FN(p,t)} = I_{FN(p,t)} + I_{NF(p,t)} \ . \qquad \text{(A-3-11)}$$

Thus, the J-divergence is a symmetrized form of I-divergence. The J-divergence satisfies several useful properties.

1. $J_{FN} > 0$ for $F \neq N$
2. $J_{FF} = J_{NN} = 0$
3. $J_{FN} = J_{NF}$
4. J_{FN} is additive for independent measurements, i.e., divergence based on t independent measurements (e.g., components of μ) is equal to the sum of the t divergences based on each measurement separately.

$$J_{FN}(\mu_1 \mu_2, \ldots, \mu_t) = \sum_{k=1}^{m} J_{FN}(\mu_k) \quad . \tag{A-3-12}$$

However, J-divergence does not satisfy the triangular inequality required of a metric:

$$J_{FN} + J_{NF'} \not> J_{FF'} \qquad \text{in general} \quad . \tag{A-3-13}$$

Finally, the Bhattacharyya distance, B_{FN}, is defined as the negative log of the Bhattacharyya coefficient, ρ_{FN},

$$\rho_{FN} = \int_{\mu} \left[p(\mu|F)\ p(\mu|N)\right]^{0.5} d\mu \tag{A-3-14}$$

$$B_{FN} = -\ell n\ \rho_{FN} \quad . \tag{A-3-15}$$

Since $\int_{\mu} p(\mu/N)d\mu = \int_{\mu} p(\mu/F)d\mu = 1$, Bhattacharyya coefficient may be regarded as the cosine of the angle between the two vectors $p(\mu/F)$ and $p(\mu/N)$ in the space spanned by the parity vector μ. Thus, ρ_{FN} is a measure of correlation between the distributions of μ under hypotheses F and N. Therefore, the Bhattacharyya coefficient, ρ_{FN} lies between 0 and 1, and the Bhattacharyya distance, B_{FN} is bounded by $0 \leq B_{FN} \leq \infty$. Note, in particular, that $\rho_{FF} = \rho_{NN} = 1$ and $B_{FF} = B_{NN} = 0$. As with the J-divergence, Bhattacharyya distance need not satisfy the triangular inequality required of a metric.

Example 3.1

Evaluate I-divergence, J-divergence and the Bhattacharyya distance for a Gaussian random vector μ of dimension t characterized by the following two distributional parameters under two hypotheses F and N.

$$p(\mu/F) = N(\bar{\mu}_F, \Sigma_F),\ p(\mu/N) = N(\bar{\mu}_N, \Sigma_N)$$

where we have assumed nonzero mean, $\bar{\mu}_N$ for μ under hypothesis N, for the sake of generality. The logarithm of the likelihood ratio, denoted by L_{FN}, is

$$L_{FN} = - \ell n \det (\Sigma_N \Sigma_F^{-1}) + \frac{1}{2} (\mu - \bar{\mu}_N)^T \Sigma_N^{-1} (\mu - \bar{\mu}_N)$$

$$- \frac{1}{2} (\mu - \bar{\mu}_F)^T \Sigma_F^{-1} (\mu - \bar{\mu}_F) \cdot \qquad \text{(A-3-16)}$$

The I-divergence is $I_{FN} = \int_\mu p(\mu/F) L_{FN} \, d\mu$ and has the form:

$$I_{FN} = - \frac{1}{2} \ell n \det (\Sigma_F \Sigma_N^{-1}) + \frac{1}{2} tr [\Sigma_F \Sigma_N^{-1} - I_t]$$

$$+ \frac{1}{2} (\bar{\mu}_F - \bar{\mu}_N)^T \Sigma_N^{-1} (\bar{\mu}_F - \bar{\mu}_N) \qquad \text{(A-3-17b)}$$

$$= - \frac{1}{2} \ell n \det (\Sigma_F \Sigma_N^{-1}) + \frac{1}{2} tr [\Sigma_F (\Sigma_N^{-1} - \Sigma_F^{-1})]$$

$$+ \frac{1}{2} (\bar{\mu}_F - \bar{\mu}_N)^T \Sigma_N^{-1} (\bar{\mu}_F - \bar{\mu}_N) \cdot \qquad \text{(A-3-17b)}$$

The J-divergence is given by $J_{FN} = I_{FN} + I_{NF}$ and has the explicit form:

$$J_{FN} = \frac{1}{2} tr [(\Sigma_F - \Sigma_N)(\Sigma_N^{-1} - \Sigma_F^{-1}) + (\Sigma_N^{-1} + \Sigma_F^{-1})(\bar{\mu}_F - \bar{\mu}_N)(\bar{\mu}_F - \bar{\mu}_N)^T] \qquad \text{(A-3-18a)}$$

$$= \frac{1}{2} tr [\Sigma_F \Sigma_N^{-1} + \Sigma_N \Sigma_F^{-1} + (\Sigma_F^{-1} + \Sigma_N^{-1})(\bar{\mu}_F - \bar{\mu}_N)(\bar{\mu}_F - \bar{\mu}_N)^T] - t \cdot \qquad \text{(A-3-18b)}$$

The Bhattacharyya distance, B_{FN} can be determined, after some tedious computations, as

$$B_{FN} = \frac{1}{2} \ell n \det [\frac{1}{2} (\Sigma_F + \Sigma_N) \Sigma_F^{-0.5} \cdot \Sigma_N^{-0.5}] + \frac{1}{8} (\bar{\mu}_F - \bar{\mu}_N) [\frac{\Sigma_F + \Sigma_N}{2}] (\bar{\mu}_F - \bar{\mu}_N) \cdot \qquad \text{(A-3-19)}$$

Two special cases are of particular interest.

Case 1. Equal covariances under two hypotheses F and N: $\Sigma_F = \Sigma_N = \Sigma$

The I-divergence in this case is symmetric and is given by

$$I_{FN} = \frac{1}{2} (\bar{\mu}_F - \bar{\mu}_N)^T \Sigma^{-1} (\bar{\mu}_F - \bar{\mu}_N) \quad . \tag{A-3-20}$$

The J-divergence is

$$J_{FN} = (\bar{\mu}_F - \bar{\mu}_N)^T \Sigma^{-1} (\bar{\mu}_F - \bar{\mu}_N) = 2\ I_{FN} \quad . \tag{A-3-21}$$

It may be noted that the term $(\bar{\mu}_F - \bar{\mu}_N)\Sigma^{-1} (\bar{\mu}_F - \bar{\mu}_N)$ is the Mahalanobis generalized distance. The Bhattacharyya distance in this case is

$$B_{FN} = \frac{1}{8} (\bar{\mu}_F - \bar{\mu}_N)\Sigma^{-1} (\bar{\mu}_F - \bar{\mu}_N) = \frac{J_{FN}}{8} \quad . \tag{A-3-22}$$

Case 2. Equal means under two hypotheses, $\bar{\mu}_F = \bar{\mu}_N$. The I-divergence, the J-divergence and the Bhattacharyya distance are given by

$$I_{FN} = -\frac{1}{2} \ln \det \left(\Sigma_F\ \Sigma_N^{-1}\right) + \frac{1}{2} \operatorname{tr} \left(\Sigma_F\ \Sigma_N^{-1}\right) - \frac{t}{2} \tag{A-3-23}$$

$$J_{FN} = \frac{1}{2} \operatorname{tr} \left[\Sigma_F\ \Sigma_N^{-1} + \Sigma_N\ \Sigma_F^{-1}\right] - t \tag{A-3-24}$$

$$B_{FN} = \frac{1}{2} \ln \det \left[\left(\Sigma_F^{0.5}\ \Sigma_N^{-0.5} + \Sigma_N^{0.5}\ \Sigma_F^{-0.5}\right)/2\right] \quad . \tag{A-3-25}$$

Note in particular that when $\| \Sigma_F\ \Sigma_N^{-1} - I_t \|$ is small with respect to any suitable norm, the I-divergence has a simpler form. In this case, note that

$$I_{FN} = -\frac{1}{2} \ell n \det \left(\Sigma_F \Sigma_N^{-1}\right) + \frac{1}{2} tr \left(\Sigma_F \Sigma_N^{-1}\right) - \frac{t}{2}$$

$$= -\frac{1}{2} \ell n \det \left((\Sigma_F \Sigma_N^{-1} - I_t) + I_t\right) + \frac{1}{2} tr \left(\left[\Sigma_F - \Sigma_N\right) \Sigma_N^{-1}\right] .$$

Expanding the ℓn term up to second order, we have

$$I_{FN} \approx -\frac{1}{2} tr \left[\Sigma_F \Sigma_N^{-1} - I_t\right] + \frac{1}{4} tr \left\{\left[\Sigma_F \Sigma_N^{-1} - I_t\right]^2\right\} + \frac{1}{2} tr \left[(\Sigma_F - \Sigma_N)\Sigma_N^{-1}\right]$$

$$\approx \frac{1}{4} tr \left[(\Sigma_F \Sigma_N^{-1} - I_t)^2\right] . \qquad \text{(A-3-26)}$$

Thus, $I_{FN} > 0$ for all $\Sigma_F \neq \Sigma_N$.

A-3.2.1 Relationship Between Distance Measures and Probability of Error

The J-divergence and the Bhattacharyya distance measures are important in their own right, but their value is enhanced by the fact that they provide bounds on the probability error (misclassification) in statistical hypothesis testing problems. In the binary hypothesis testing problem, let $\mu^k = (\mu(1)$ $\mu(2),\ldots,\mu(k))$ be the time-sequence of parity checks, each parity check $\mu(i)$ being a t dimensional vector. Let $p(\mu^k/F)$ and $p(\mu^k/N)$ be the probability density functions of μ^k under hypotheses F and N, respectively; and δ_F, $\delta_N(=1-\delta_F)$ be the a priori probabilities (based on mean-time-between-failures of components) that the system is failed and normal, respectively. Then, the probability of error of the maximum a posteriori (MAP) decision rule is $P_e(k)$ and is given by

$$P_e(k) = \int_{\mu^k} \min \left(\delta_F\, p(\mu^k/F),\ \delta_N\, p(\mu^k/N)\right) d\mu^k \qquad \text{(A-3-27)}$$

The multi-dimensional integral in Eq. A-3-27 is extremely difficult to evaluate, even with numerical integration. For this reason, the J-divergence and the Bhattacharyya distance are used in signal selection problems of communication theory and in feature selection problems of pattern recognition, because they provide the following bounds on $P_e(k)$ [32-34].*

$$0.5 - 0.5 \left(1 - 4\ \delta_N \delta_F\ \rho_{FN}^2(k)\right)^{0.5} \leq P_e(k) \leq \left(\delta_N \delta_F\right)^{0.5}\ \rho_{FN}(k) \qquad \text{(A-3-28)}$$

$$0.5 \min\left(\delta_N \delta_F\right) \exp\left(-J_{FN}(k)/8\right) \leq P_e(k) \leq \left(\delta_N \delta_F\right)^{0.5}\ \left(J_{FN}(k)/4\right)^{-0.25} \qquad \text{(A-3-29)}$$

with upper bound in Eq. A-3-29 valid only for Gaussian statistics. In Eqs. A-3-28 and A-3-29, $\rho_{FN}(k)$ is the Bhattacharyya coefficient and $J_{FN}(k)$ is the J-divergence based on μ^k:

$$\rho_{FN}(k) \overset{\Delta}{=} \int_{\mu^k} \left(p(\mu^k/F) \cdot p(\mu^k/N)\right)^{0.5}\ d\mu^k \qquad \text{(A-3-30)}$$

$$J_{FN} \overset{\Delta}{=} \int_{\mu^k} \left(p(\mu^k/F) - p(\mu^k/N)\right) \ln \frac{p(\mu^k/F)}{p(\mu^k/N)}\ d\mu^k \ . \qquad \text{(A-3-31)}$$

When the signal to noise ratio is high (i.e., as $P_e \to 0$), it is stated in Ref. [32] that

$$\ln P_e(k) \approx \ln \rho_{FN}(k) = - B_{FN}(k) \text{ as } P_e(k) \to 0 \qquad \text{(A-3-32)}$$

where $B_{FN}(k)$ is the Bhattacharyya distance between the probability distributions $p(\mu^k/F)$ and $p(\mu^k/N)$.

*Note that, in the subsection, we have suppressed the dependence of ρ, J and I on the order of parity relations, p and the number of parity relations, t for convenience of notation.

A-3.2.2 Asymptotic per Sample Distance Measures

As the number of samples of parity checks, $k \to \infty$, the asymptotic per sample I-divergence is

$$\bar{I}_{FN} = \lim_{k \to \infty} \frac{1}{k} I_{FN}(k) \tag{A-3-33}$$

where

$$I_{FN}(k) = \int_{\mu^k} p(\mu^k/F) \ln \frac{p(\mu^k/F)}{p(\mu^k/N)} \cdot d\mu^k \quad . \tag{A-3-34}$$

It is shown in Ref. [33] that $\bar{I}_{FN}$ is given by

$$\bar{I}_{FN} = (4\pi)^{-1} \int_0^{2\pi} \{ \mathrm{tr}[\Phi_{\mu N}^{-1}(\omega) \Phi_{\mu F}(\omega) - I_t] - \ln \det (\Phi_{\mu N}^{-1}(\omega) \Phi_{\mu F}(\omega))$$

$$+ (\bar{\mu}_F(\omega) - \bar{\mu}_N(\omega))^T \Phi_{\mu N}^{-1}(\omega) (\bar{\mu}_F(\omega) - \bar{\mu}_N(\omega)) \} d\omega \tag{A-3-35}$$

where $\Phi_{\mu F}(\omega)$ and $\Phi_{\mu N}(\omega)$ are the spectral density matrices of residuals under the failed and normal hypotheses, respectively. In addition, $\bar{\mu}_F(\omega)$ are the discrete Fourier transforms of $\bar{\mu}_F(k)$ and $\bar{\mu}_N(k)$, respectively. The asymptotic per sample J-divergence is evaluated as $\bar{J}_{FN} = \bar{I}_{FN} + \bar{I}_{NF}$. Similarly, the asymptotic per sample Bhattacharyya distance is given by

$$\bar{B}_{FN} = (4\pi)^{-1} \int_0^{2\pi} \{ \ln \det [\Phi_{\mu F}^{-0.5}(\omega) \Phi_{\mu N}^{0.5}(\omega)$$

$$+ \Phi_{\mu F}^{0.5}(\omega) \Phi_{\mu N}^{-0.5}(\omega)/2)] + \frac{1}{4} (\bar{\mu}_F(\omega) - \bar{\mu}_N(\omega))^T [(\Phi_{\mu F}(\omega) \tag{A-3-36}$$

$$+ \Phi_{\mu N}(\omega)/2)]^{-1} (\bar{\mu}_F(\omega) - \bar{\mu}_N(\omega)) \} d\omega$$

Interestingly, Eqs. A-3-35 and A-3-36 are valid even if the deterministic mean components $\bar{\mu}_N$ and $\bar{\mu}_F$ are functions of frequency, ω. This generality allows us to consider a variety of bias, ramp and other periodic failure modes. However, asymptotic measures are of limited utility in FDI, where one is interested in the transient effects of failure modes.

A-3.2.3 Relation to Fisher Information Matrix (FIM):

In parameter estimation problems (i.e., a continuum of hypotheses), it turns out that the I-and J-divergence measures, as well as the Bhattacharyya coefficient, ρ (or equivalently, the Bhattacharyya distance, B) are related to the Fisher information matrix (FIM). The use of FIM for optimal input design for system identification, and for the design of insensitive feedback control laws is well known [42],[43]. This relationship provides additional motivation for the use of divergence measures and the Bhattacharyya distance measures for parity generation. To show this relationship, consider the Bhattacharyya coefficient between $p(\mu^k/\theta)$ and $P(\mu^k/\theta+\Delta\theta)$, denoted by $\rho_{\theta,\theta+\Delta\theta}$ is:

$$\rho_{\theta,\theta+\Delta\theta} = \int_{\mu^k} \left(p(\mu^k/\theta) \cdot p(\mu^k/\theta+\Delta\theta)\right)^{0.5} d\mu^k \qquad \text{(A-3-37)}$$

The Fisher information matrix, F_θ, on the other hand, is given by

$$F_\theta = \int_{\mu^k} \left(\frac{\partial}{\partial\theta} \ell n\ p(\mu^k/\theta)\right) \left(\frac{\partial}{\partial\theta} \ell n\ p(\mu^k/\theta)\right)^T p(\mu^k/\theta)\ d\mu^k \qquad \text{(A-3-38a)}$$

$$= \int_{\mu^k} \left(\frac{\partial}{\partial\theta} p(\mu^k/\theta)\right) \left(\frac{\partial}{\partial\theta} p(\mu^k/\theta)\right)^T p(\mu^k/\theta)^{-1} d\mu^k \qquad \text{(A-3-38b)}$$

To exhibit the relationship between Eqs. A-3-37 and A-3-38, we expand $p(\mu^k|\theta+\Delta\theta)$ to second order using Taylor series:

$$p(\mu^k/\theta+\Delta\theta) \cong p(\mu^k/\theta) + \Delta\theta^T \cdot \frac{\partial}{\partial\theta} p(\mu^k/\theta) + \frac{1}{2} \Delta\theta^T \frac{\partial^2}{\partial\theta^2} p(\mu^k/\theta) + o(\|\Delta\theta\|^3) \quad \text{(A-3-39)}$$

where $\frac{\partial^2}{\partial\theta^2} p(\mu^k/\theta)$ is the Hessian of $p(\mu^k/\theta)$ with respect to θ and $\|$. $\|$ is the Euclidean norm. Using Eq. A-3-39, we can evaluate $\rho_{\theta,\theta+\Delta\theta}$ to second order as

$$\rho_{\theta,\theta+\Delta\theta} \cong \int_{\mu k} \left[p(\mu^k/\theta) + \frac{1}{2} \Delta\theta^T \frac{\partial}{\partial\theta} p(\mu^k/\theta) + \Delta\theta^T \frac{\partial^2}{\partial\theta^2} p(\mu^k/\theta)\Delta\theta \right.$$

$$\left. - \frac{1}{8} \Delta\theta^T \left(\frac{\partial p}{\partial\theta} p(\mu^k/\theta)\right) \left(\frac{\partial}{\partial\theta} p(\mu^k/\theta)\right)^T \left(p(\mu^k/\theta)\right)^{-1} \Delta\theta\right] d\mu^k \quad \text{(A-3-40)}$$

Note that $\int_{\mu^k} p(\mu^k/\theta)d\mu^k = 1$, and also the second and third terms integrate to zero, and Eq. A-3-40 reduces to

$$\rho_{\theta,\theta+\Delta\theta} = 1 - \frac{1}{8} \Delta\theta^T F_\theta \Delta\theta \quad \text{(A-3-41)}$$

Thus, in a local sense, we have the following: small Bhattacharyya coefficient, ρ implies larger Bhattacharyya distance, which in turn, implies larger information content in μ^k to distinguish (or identify) θ.

Similarly, one can show that the I-divergence, for small parameter variations, is given by

$$I_{\theta,\theta+\Delta\theta} = \frac{1}{2} \Delta\theta^T F_\theta \Delta\theta \quad \text{(A-3-42)}$$

and the J-divergence is given by

$$J_{\theta,\theta+\Delta\theta} = \Delta\theta^T F_\theta \Delta\theta \quad \text{(A-3-43)}$$

The relations (Eqs. A-3-41 - A-3-43) imply that $1-\rho$, I and J satisfy metric properties of topology locally (but not globally). They also imply that if the two hypotheses F and N differ by a small amount, the Fisher information matrix provides a convenient measure of distinguishability.

A-3.2.4 Distinguishability of Hypotheses (Failure Modes)

In the zero mean case, i.e, $\bar{\mu}_F(\omega) = \bar{\mu}_N(\omega) = 0$, in Eq. A-3-36, it is shown in Ref. [34] that the Bhattacharyya coefficient $\rho_{FN}(k) \to 0$ (and, consequently, the probability of error, $P_e(k) \to 0$) as long as the asymptotic per sample Bhattacharyya distance $\bar{B}_{FN}$ of Eq. A-3-36 is nonzero. A necessary and sufficient condition for $\bar{B}_{FN} = 0$ is that the eigenvalues λ_i, $i = 1,2,\ldots,t$ of the generalized eigenvalue equation:

$$\Phi_{\mu F}(\omega)\, g_i = \lambda_i\, \Phi_{\mu N}(\omega)\, g_i \;;\; i=1,2,\cdots,t \qquad \text{(A-3-44)}$$

satisfy

$$\lambda_i(\omega) = 1 \text{ almost everywhere on } \omega \in [0,2\pi] \quad . \qquad \text{(A-3-45)}$$

Equation A-3-44 provides a simple check on the "discriminability" (or distinguishability) of two hypotheses for a given residual generation mechanism. Looked another way, it also tells us that as long as the Bhattacharyya distance, $\bar{B}_{FN}$ is nonzero, one can devise parity relations that achieve discrimination between two hypotheses. However, this analysis does not address the issue of speed of such discrimination, which is discussed next.

A-3.2.5 Speed of Discrimination

It is also shown in Ref. [34] that, for stationary dynamic models, the Bhattacharyya coefficient $\rho_{FN}(k)$ (and hence $P_e(K)$) tends to zero exponentially with the number of observations, k, and that the rate of covergence is governed by the asymptotic per sample Bhattacharyya distance, $\bar{B}_{FN}$, as

$$\exp\{-\alpha(k) - k\,\bar{B}_{FN}\} \leq \rho_{FN}(k) \leq \exp\{\alpha(k) - k\,\bar{B}_{FN}\} \qquad \text{(A-3-46)}$$

where $\alpha(k)$ is any function that satisfies $\lim_{k\to\infty} k^{-1}\ \alpha(k) \to 0$ and $\bar{B}_{FN}$ was defined earlier in Eq. A-3-36. Since $\rho_{FN}(k)$ provides an upper bound on the probability of error (see Eq. A-3-38), we can compute a lower bound on the detection delay, k_d. For a desired probability of error, P_{edes}

$$k_d \geq - \ell n\left[P_{edes} \cdot \left(\delta_N \delta_F\right)^{-0.5}\right]/\bar{B}_{FN} \quad . \tag{A-3-47}$$

In addition, Eq. A-3-46 says that a parity generation mechanism that has larger asymptotic per sample Bhattacharyya distance $\bar{B}_{FN}$ is likely to be an exponen- tially faster discriminator of failed and normal modes than one with a smaller distance. Second, when a decision statistic is being designed, the bounds in Eqs. A-3-28 and A-3-29 provide a means to select the required data collection window for a given probability of error. Specifically, if we view the data collection process as part of the parity check generation process, i.e., we are generating parity checks of relatively high-order p, then p should chosen so that

$$\rho_{FN}(p,t) \approx \left(\delta_N \delta_F\right)^{-0.5} \cdot P_{edes} \tag{A-3-48}$$

where we have shown the explicit dependence of the Bhattacharyya coefficient, ρ on the order and number of parity relations p and t.

A-3.2.6 Divergence and the Bhattacharyya Distance in Uncertain Models

When the model parameters are uncertain as in Eqs. A-3-1 through A-3-4, we define the I-divergence, $I_{FN}(p,t)$, as the average discriminating information for failure hypothesis F with respect to hypothesis N over all possible models, $\ell = 1,2,\cdots,L$. That is,

$$I_{FN}(p,t) \stackrel{\Delta}{=} \sum_{\ell=1}^{L} P_\ell E_\ell \left\{I_{FN}(p,t)/\ell\right\} \tag{A-3-49}$$

where $I_{FN/\ell}(p,t)$ is the I-divergence between hypotheses F and N conditioned on model ℓ. Letting $\Sigma_{F\ell}$ and $\Sigma_{N\ell}$ be the steady-state covariances of $\mu(k)$ under hypotheses F and N, under model ℓ, and $\bar{\mu}_{F\ell}$ and $\bar{\mu}_{N\ell}$ be the corresponding means, we can compute $I_{FN/\ell}(p,t)$ from Eq. A-3-17 as:

$$I_{FN/\ell}(p,t) = \frac{1}{2}\left[-\ell n \det\left(\Sigma_{F\ell}\,\Sigma_{N\ell}^{-1}\right) + \mathrm{tr}\left(\Sigma_{F\ell}\,\Sigma_{N\ell}^{-1} - I_t\right)\right.$$

$$\left. + \left(\bar{\mu}_{F\ell} - \bar{\mu}_{N\ell}\right)^T \Sigma_{N\ell}^{-1}\left(\bar{\mu}_{F\ell} - \bar{\mu}_{N\ell}\right)\right] \qquad \text{(A-3-50)}$$

Similarly, the J-divergence is given by

$$J_{FN}(p,t) \triangleq \sum_{\ell=1}^{L} P_\ell E_\ell \left\{J_{FN/\ell}(p,t)\right\} \qquad \text{(A-3-51)}$$

where $J_{FN/\ell}(p,t)$ is obtained from Eq. A-3-18 as

$$J_{FN/\ell}(p,t) = \frac{1}{2}\,\mathrm{tr}\left[\left(\Sigma_{F\ell} - \Sigma_{N\ell}\right)\left(\Sigma_{F\ell}^{-1} - \Sigma_{N\ell}^{-1}\right)\right]$$

$$+ \frac{1}{2}\left(\bar{\mu}_{F\ell} - \bar{\mu}_{N\ell}\right)^T \left(\Sigma_{F\ell}^{-1} + \Sigma_{N\ell}^{-1}\right)\left(\bar{\mu}_{F\ell} - \bar{\mu}_{N\ell}\right) \qquad \text{(A-3-52)}$$

In a similar vein, the Bhattacharyya coefficient $\rho_{FN}(p,t)$ is given by

$$\rho_{FN}(p,t) = \sum_{\ell=1}^{L} P_\ell\, \rho_{FN/\ell}(p,t) \qquad \text{(A-3-53)}$$

where $\rho_{FN/\ell}(p,t)$ is given by [22]

$$\rho_{FN/\ell}(p,t) = \det\left[\Sigma_{F\ell}^{0.5}\,\Sigma_{N\ell}^{-0.5} + \Sigma_{F\ell}^{-0.5}\,\Sigma_{N\ell}^{0.5}/2\right] \cdot \exp\left\{-\frac{1}{8}\left(\bar{\mu}_{F\ell} - \bar{\mu}_{N\ell}\right)^T\right.$$

$$\left.\left(\Sigma_{F\ell} + \Sigma_{N\ell}/2\right)^{-1}\left(\bar{\mu}_{F\ell} - \bar{\mu}_{N\ell}\right)\right. \qquad \text{(A-3-54)}$$

A closed form expression for $B_{FN}(p,t) = -\ \ell n\ \rho_{FN}(p,t)$ is not possible, since $\rho_{FN}(p,t)$ in Eq. A-3-53 is a sum of exponential terms. However, since $-\ \ell n\ x$ is a convex function of x, we can use Jensen's inequality [22] to obtain an upper bound on $B_{FN}(p,t)$:

$$B_{FN}(p,t) < \sum_{\ell=1}^{L} P_\ell\ B_{FN/\ell}(p,t) \qquad \text{(A-3-55)}$$

where $B_{FN/\ell}(p,t)$ is obtained from Eq. A-3-19 as

$$B_{FN/\ell}(p,t) = \frac{1}{2}\ \ell n\ \det\ [\Sigma_{F\ell}^{0.5}\ \Sigma_{N\ell}^{-0.5} + \Sigma_{F\ell}^{-0.5}\ \Sigma_{N\ell}^{0.5}/2]$$

$$+ \frac{1}{8}\ (\bar{\mu}_{F\ell} - \bar{\mu}_{N\ell})^T\ (\Sigma_{F\ell} + \Sigma_{N\ell}\ /2)^{-1}\ (\bar{\mu}_{F\ell} - \bar{\mu}_{N\ell}) \qquad \text{(A-3-56)}$$

If we make the Gaussian sum approximation in Eqs. A-3-49 - A-3-56, then the I-divergence $I_{FN}(p,t)$, the J-divergence $J_{FN}(p,t)$ and the Bhattacharyya distance $B_{FN}(p,t)$ (approximately) reduce to Eqs. A-3-17 -A-3-19, respectively, with the following identifications*

$$\Sigma_F = \sum_{\ell=1}^{L} P_\ell\ \Sigma_{F\ell} \qquad \text{(A-3-57a)}$$

$$\Sigma_N = \sum_{\ell=1}^{L} P_\ell\ \Sigma_{N\ell} \qquad \text{(A-3-57b)}$$

$$\bar{\mu}_F = \sum_{\ell=1}^{L} P_\ell\ \bar{\mu}_{F\ell} \qquad \text{(A-3-57c)}$$

$$\bar{\mu}_N = \sum_{\ell=1}^{L} P_\ell\ \bar{\mu}_{N\ell} \qquad \text{(A-3-57d)}$$

*The approximation stems from the neglect of mean related term in the covariance of the Gaussian sum approximation, i.e., Eqs. A-3-57a and A-3-57b.

A-3.2.7 Extension to Multiple Hypotheses

If there are more than two hypotheses $H \in \mathcal{H}$ (e.g., $H_0 = N$, $H_1 = F$, $H_2 = F'$ etc.), then it appears that the following average J-divergence and the Bhattacharyya coefficients are suitable measures:

$$J(p,t) = \sum_{H \in \mathcal{H}} \sum_{H' \in \mathcal{H}} \delta_H \, \delta_{H'} \, J_{HH'}(p,t) \qquad \text{(A-3-58)}$$

$$\rho(p,t) = \sum_{H \in \mathcal{H}} \sum_{H' \in \mathcal{H}} \delta_H \, \delta_{H'} \, \rho_{HH'}(p,t) \; . \qquad \text{(A-3-59)}$$

In the multiple hypothesis case, it is known that the total error probability P_{etot} and the pair-wise error between hypotheses H and H', denoted by $P_{eHH'}$, given by Eq. A-3-27 are related by the inequality [44]

$$P_{etot} \leq \frac{1}{2} \sum_{H \in \mathcal{H}} \sum_{H' \in \mathcal{H}} P_{eHH'} \; . \qquad \text{(A-3-60)}$$

Since the Bhattacharyya coefficient $\rho_{HH'}$ provides an upper bound on the probability of error (see Eq. A-3-28), we have

$$P_{etot} \leq \frac{1}{2} \sum_{H \in \mathcal{H}} \sum_{H' \in \mathcal{H}} \delta_H \, \delta_{H'} \, \rho_{HH'} \; . \qquad \text{(A-3-61)}$$

In addition, since $\delta_H \delta_{H'} \leq \frac{1}{2}$, we have

$$P_{etot} \leq \frac{1}{4} \sum_{H \in \mathcal{H}} \sum_{H' \in \mathcal{H}} \rho_{HH'} \; . \qquad \text{(A-3-62)}$$

The average measures have limited practical utility in our decentralized FDI approach. We use pairwise measures to generate the multilevel parity generation scheme discussed in Section A-1 and subsection A-3.1.

A-3.3 PARITY GENERATION TO MAXIMIZE THE DIVERGENCE MEASURES AND BHATTACHARYYA DISTANCE

In this subsection, we use the divergence and the Bhattacharyya distance based measures to generate parity relations for the system described by Eqs. A-3-1 through A-3-4 to detect a failure. The key idea here is to generate the parity transformation matrix G that maximizes separation between normal and failure modes of system operation, i.e., to find parity checks that are representative of the dissimilarities between normal and failure modes.

Using the Gaussian sum approximation embodied in Eq. A-2-60; letting $\hat{C}_N$ and $\hat{C}_F$ denote the average ACF counterparts of Eq. A-2-60, and $\bar{Y}_{pF}$ denote the mean of Y_p under failure hypothesis (note that $\bar{Y}_{pN} = 0$ by assumption), we have

$$\Sigma_F = G^T \hat{C}_F G \tag{A-3-63a}$$

$$\Sigma_N = G^T \hat{C}_F G \tag{A-3-63b}$$

$$\bar{\mu}_F = G^T \bar{Y}_{pF} \tag{A-3-63c}$$

$$\bar{\mu}_N = G^T \bar{Y}_{pN} = 0 \tag{A-3-63d}$$

Assuming, without loss of generality, that the failure occurred at k-p, $\bar{Y}_{pF}$ is given by

$$\bar{Y}_{pF} = \Gamma_p \, d_{pF}(k) + b_{pF}(k) \tag{A-3-64}$$

where Γ_p can be evaluated directly from the system matrices $\{C_\ell, A_\ell\}$ and prior probabilities, P_ℓ as:

$$\Gamma = \sum_{\ell=1}^{L} P_\ell \, \Gamma_{p\ell} \tag{A-3-65a}$$

$$\Gamma_{p\ell} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ C_{F\ell} & 0 & \cdots & 0 \\ C_{F\ell}A_{F\ell} & C_{F\ell} \; 0 & \cdots & 0 \\ \vdots & \vdots & & \\ C_{F\ell}A_{F\ell}^{p-1} & \cdots & C_{F\ell} & 0 \end{bmatrix} \tag{A-3-65b}$$

$$d_{pF}^T(k) \triangleq [d(k-p) \;\; d(k-p+1) \;\; \cdots \;\; d(k)] \;\; \text{a } (p+1)n \text{ vector}$$

$$b_{pF}^T(k) \triangleq [b(k-p) \;\; b(k-p+1) \;\; \cdots \;\; b(k)] \;\; \text{a } (p+1)m \text{ vector} \; .$$

We can explicitly evaluate the divergence measures and the Bhattacharyya distance (Eqs. A-3-17 and A-3-19) in terms of G, $\hat{C}_N$, $\hat{C}_F$ and $\bar{Y}_{pF}$ as:

$$I_{FN}(p,t) = \frac{1}{2} \bar{Y}_{pF}^T \, G \, (G^T\hat{C}_N G)^{-1} G^T \bar{Y}_{pF} - \frac{1}{2} \ln \det [(G^T\hat{C}_F G)(G^T\hat{C}_N G)^{-1}]$$

$$+ \frac{1}{2} \operatorname{tr} [(G^T\hat{C}_F G)(G^T\hat{C}_N G)^{-1}] \tag{A-3-66}$$

$$J_{FN}(p,t) = \frac{1}{2} \bar{Y}_{pF}^T \, G \, [(G^T\hat{C}_N G)^{-1} + (G^T\hat{C}_F G)^{-1}] G^T \bar{Y}_{pF}$$

$$+ \frac{1}{2} \operatorname{tr} [(G^T\hat{C}_N G)^{-1}(G^T\hat{C}_F G) + (G^T\hat{C}_F G)^{-1}(G^T\hat{C}_N G) - 2I_t] \tag{A-3-67}$$

$$B_{FN}(p,t) = \frac{1}{2} \ell n \det \left[\left(G^T \hat{C}_N G\right)^{-0.5} \left(G^T \hat{C}_F G\right)^{0.5} \right.$$

$$\left. + \left(G^T \hat{C}_F G\right)^{-0.5} \left(G^T \hat{C}_N G\right)^{0.5} / 2 \right]$$

$$+ \frac{1}{8} \bar{Y}_{pF}^T G \left[G^T \left(\hat{C}_N + \hat{C}_F\right) G/2\right]^{-1} G^T \bar{Y}_{pF} \qquad \text{(A-3-68)}$$

The general optimization approach for all distance measures involves a gradient-type scheme. However, in two special cases, which are of considerable interest in the FDI problem, the optimization provides explicit solutions. The general problem can also be solved suboptimally. We will discuss these next.

A-3.3.1 Detection of Bias Failures of Known Magnitude

In this case $A_{N\ell} = A_{F\ell}$, $D_{N\ell} = D_{F\ell}$ for all ℓ. This implies that the average ACFs $\hat{C}_F \cong \hat{C}_N = \hat{C}^*$, and the distance measures (Eqs. A-3-66 - A-3-68) are scaled versions of the Mahalanobis distance [17]:

$$I_{FN}(p,t) = \frac{1}{2} J_{FN}(p,t) = 4 B_{FN}(p,t) = \frac{1}{2} \bar{Y}_{pF}^T G \left(G^T \hat{C} G\right)^{-1} G^T \bar{Y}_{pF} \qquad \text{(A-3-69)}$$

The optimal parity relation to maximize the criterion in Eq. A-3-69 is given by the following theorem.

*The approximation $C_F \cong C_N$ comes from neglecting the mean related term in the covariance of the Gaussian sum approximation.

Theorem 3.1: There exists a single best parity relation (i.e., t=1) that maximizes Eq. A-3-69 and is given by

$$\mu_1(k) = g_1^T Y_p(k) = \bar{Y}_{pF}^T \hat{C}^{-1} Y_p(k) \quad . \tag{A-3-70}$$

The parity transformation G^* is

$$G^* = g_1 = \hat{C}^{-1} \bar{Y}_{pF} \tag{A-3-71}$$

and the optimal distance measure is

$$I_{FN}^*(p,t) = \frac{1}{2} J_{FN}^*(p,t) = 4 B_{FN}^*(p,t) = \frac{1}{2} \bar{Y}_{pF}^T \hat{C}^{-1} \bar{Y}_{pF} \quad . \tag{A-3-72}$$

Proof: The I-divergence measure $I_{FN}(p,t)$ in Eq. A-3-69 can be written as a trace functional

$$I_{FN}(p,t) = \frac{1}{2} \text{tr} \left[(G^T \hat{C} G)^{-1} G^T \bar{Y}_{pF} \bar{Y}_{pF}^T G \right] \quad . \tag{A-3-79}$$

The optimal G^* satisfies the gradient relationship:

$$0 = \nabla_G I_{FN} = \bar{Y}_{pF} \bar{Y}_{pF}^T G^* (G^{*T} \hat{C} G^*)^{-1}$$

$$- \hat{C} G^* (G^{*T} \hat{C} G^*)^{-1} G^{*T} \bar{Y}_{pF} \bar{Y}_{pF}^T G^* (G^{*T} \hat{C} G^*)^{-1} = 0 \quad . \tag{A-3-74}$$

Pre-multiplying and post-multiplying Eq. A-3-74 by $\hat{C}^{-1}$ and $G^{*T} \hat{C} G^*$, we have the simplified relationship:

$$\hat{C}^{-1} \bar{Y}_{pF} \bar{Y}_{pF}^T G^* = G^* (G^{*T} \hat{C} G^*)^{-1} G^{*T} \bar{Y}_{pF} \bar{Y}_{pF}^T G^* \quad . \tag{A-3-75}$$

It is well known [22] the the product $\left(G^{*T}\ \hat{C}\ G^{*}\right)^{-1}\ G^{*T}\ \bar{Y}_{pF}\ \bar{Y}_{pF}^{T}\ G^{*}$ can be diagonalized by an orthonormal matrix Ψ such that

$$\left(G^{*T}\ \hat{C}\ G^{*}\right)^{-1}\ G^{*T}\ \bar{Y}_{pF}\ \bar{Y}_{pF}^{T}\ G^{*}\ \Psi = \Psi\ \Lambda_{t}\ . \qquad \text{(A-3-76)}$$

Using Eq. A-3-76 in Eq. A-3-75, we have

$$\hat{C}^{-1}\ \bar{Y}_{pF}\ \bar{Y}_{pF}^{T}\ G^{*}\ \Psi = G^{*}\ \Psi\ \Lambda_{t}\ . \qquad \text{(A-3-77)}$$

Since a t by t orthonormal matrix does not change I_{FN}, we can include Ψ in G^{*}. Since Eq. A-3-77 is an eigenvalue equation, we conclude that the optimum parity transformation G^{*} is achieved by selecting the t eigenvectors corresponding to the largest eigenvalues of $\hat{C}^{-1}\ \bar{Y}_{pF}\ \bar{Y}_{pF}^{T}$. However, the rank of $\hat{C}^{-1}\ \bar{Y}_{pF}\ \bar{Y}_{pF}^{T}$ is one, therefore, we obtain a single best parity relation (i.e., t=1). Indeed, the largest eigenvalue is

$$\lambda_{max} = \bar{Y}_{pF}^{T}\ \hat{C}^{-1}\ \bar{Y}_{pF}\ . \qquad \text{(A-3-78)}$$

Equation A-3-78 is obtained by noting the fact that rank of $\hat{C}^{-1}\ \bar{Y}_{pF}\ \bar{Y}_{pF}^{T}$ is one (i.e., it has a single non-zero eigenvalue), and that the trace of a matrix is the sum of its eigenvalues:

$$\text{tr}\ \left(\hat{C}^{-1}\ \bar{Y}_{pF}\ \bar{Y}_{pF}^{T}\right) = \bar{Y}_{pF}^{T}\ \hat{C}^{-1}\ \bar{Y}_{pF}\ . \qquad \text{(A-3-79)}$$

Using Eq. A-3-78 in the eigenvalue equation

$$\hat{C}^{-1}\ \bar{Y}_{pF}\ \bar{Y}_{pF}^{T}\ g_{1} = \bar{Y}_{pF}^{T}\ \hat{C}^{-1}\ \bar{Y}_{pF}\ g_{1}\ . \qquad \text{(A-3-80)}$$

Clearly, the optimal g_l is given by Eq. A-3-71. The optimal distance measure in Eq. A-3-72 is obtained by substituting Eq. A-3-71 in Eq. A-3-69.

The parity relation $\mu_l(k)$ can be thought of as an approximate whitener followed by a correlator, as shown in Fig. A-3-1. The approximation stems from the Gaussian sum approximation in Eq. A-3-60. The parity relation (Eq. A-3-70) is precisely the decision statistic obtained when we consider the problem of detecting known signals with (p+1) observations in a zero-mean colored noise with autocovariance function $\hat{C}$ and signal mean $\bar{Y}_{pF}$. To show this relationship, consider an observation model:

$$z(r) = s_q(r) + n(r) \quad ; \quad q = 1,2, 0 \leq r \leq p \tag{A-3-81}$$

where n(r) is a zero mean Gaussian noise process. The autocorrelation function of $\left(n(0)\cdots n(p)\right)$ is $\hat{C}$. We let

$$s_1(r) = 0 \tag{A-3-82}$$

$$s_2(r) = \bar{y}_F\,(k-p+r) \tag{A-3-83}$$

where $\bar{y}_F(k-p+r)$ is the average output over all models $\ell=1,2,\cdots,L$ under failure hypothesis. Then, the optimum likelihood test is to decide s_1 if and only if:

$$\left(S_{p1}-S_{p2}\right)^T \hat{C}^{-1} Z_p \quad > \quad \eta + \frac{1}{2}\left(S_{p1}^T \hat{C}^{-1} S_{p1} - S_{p2}^T \hat{C}^{-1} S_{p2}\right) \tag{A-3-84}$$

where η is the threshold that is a function of prior probabilities, and costs associated with false alarms and missed detections. S_{p1}, S_{p2} and Z_p are extended observations of length p, defined by

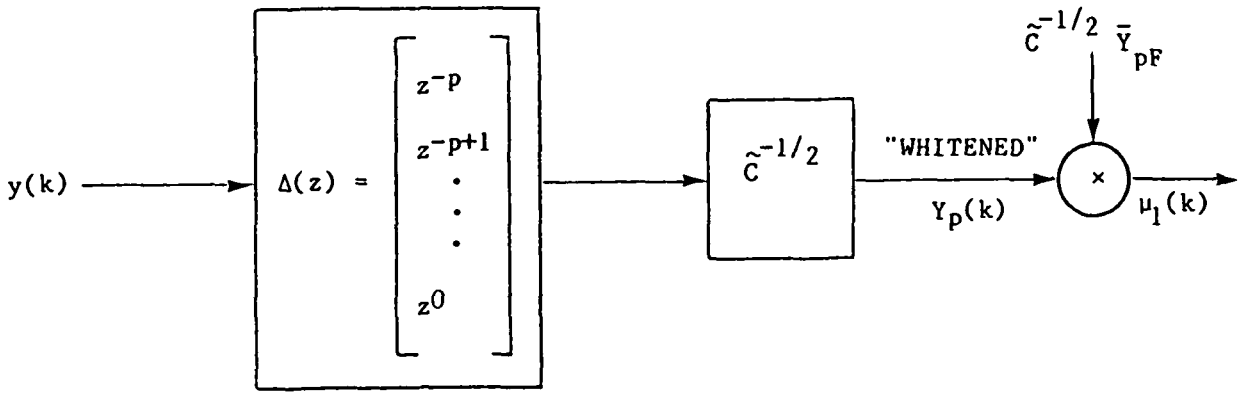


Figure A-3-1. Whitener-Correlator Interpretation of Detection Parity Relation

$$Z_p^T \triangleq [z(0) \quad z(1) \quad \cdots \quad z(p)] = Y_p^T(k) \tag{A-3-85a}$$

$$S_{p1}^T \triangleq [0 \quad 0 \quad \cdots \quad 0] \tag{A-3-85b}$$

$$S_{p2}^T \triangleq [\bar{y}_F\ (k-p+r) \quad \cdots \quad \bar{y}_F(k)] = \bar{Y}_{pF}^T \quad . \tag{A-3-85c}$$

The left hand side of Eq. A-3-84 is precisely the information collection phase that we alluded to in Section 1. Thus $\mu_1(k)$ obtained from the divergence and the Bhattacharyya distance based measures is already accomplishing the parity generation and information collection phases of the FDI process.

A-3.3.2 Detection of Noise Variance Changes and Scale Factor Failures

This case corresponds to the situation where the noise processes have different covariances under all failure modes and uncertain models, but the additive disturbances, $d_F(k)$, and the bias term, $b_F(k)$, are zero. As a result, $\bar{Y}_{pF}(k) = 0$ and $\hat{C}_N \neq \hat{C}_F$. With this simplification, maximization of the J-divergence and the Bhattacharyya distance yield identical optimal parity

transformations G, while the transformation that maximizes the I-divergence is different. The optimal G is given by the following theorem.

Theorem 3.2: The J-divergence and the Bhattacharyya distance are maximized if the columns of G are chosen to be the t eigenvectors of $\hat{C}_N^{-1}\hat{C}_F$ corresponding to the eigenvalues λ_r for which

$$\lambda_r+\lambda_r^{-1} \geq \lambda_q+\lambda_q^{-1} \quad ; \quad q=1,2,\cdots,(p+1)m \quad ; \quad r=1,2,\cdots,t \tag{A-3-86}$$

while the I-divergence is maximized if the columns of G are chosen to be the t normalized eigenvectors associated with the t largest eigenvalues of $\hat{C}_N^{-1}\hat{C}_F$.

The optimal distance measures for the selected parity relations are:

$$J_{FN}^{*}(p,t) = \frac{1}{2}\sum_{r=1}^{t}\left(\lambda_r+\lambda_r^{-1}-2\right) \tag{A-3-87a}$$

$$B_{FN}^{*}(p,t) = \frac{1}{2}\sum_{r=1}^{t} \ell n \left(\lambda_r^{0.5} + \lambda_r^{-0.5}\right)/2 \tag{A-3-87b}$$

$$I_{FN}^{*}(p,t) = \frac{1}{2}\sum_{r=1}^{t}\left(\lambda_r - \ell n\ \lambda_r - 1\right) \quad . \tag{A-3-87c}$$

Proof: We will consider each measure, in turn.

a. Maximization of J-Divergence

The gradient expression for the J-divergence of Eq. A-3-67 when $\bar{Y}_{pF}=0$ is given by

$$\begin{aligned}\nabla_G J_{FN}(p,t) = &- \hat{C}_N G \left(G^T \hat{C}_N G\right)^{-1}\left(G^T \hat{C}_F G\right)\left(G^T \hat{C}_N G\right)^{-1} \\ &- \hat{C}_F G \left(G^T \hat{C}_F G\right)^{-1}\left(G^T \hat{C}_N G\right)\left(G^T \hat{C}_F G\right)^{-1} \\ &+ \hat{C}_F G \left(G^T \hat{C}_N G\right)^{-1} + \hat{C}_N G \left(G^T \hat{C}_F G\right)^{-1} \quad .\end{aligned} \tag{A-3-88}$$

Pre-multiplying Eq. A-3-88 by $\hat{C}_N^{-1}$, post-multiplying by $(G^T\hat{C}_N G)$, and setting the gradient to zero, the optimum G^* satisfies the relationship

$$\hat{C}_N^{-1}\ \hat{C}_F\ G^* = G^*\ (G^{*T}\ \hat{C}_N\ G^*)^{-1}\ (G^{*T}\ \hat{C}_F\ G^*)\quad . \tag{A-3-89}$$

As before, $(G^{*T}\ \hat{C}_N\ G^*)^{-1}\ (G^{*T}\ \hat{C}_F\ G^*)$ can be diagonalized via an orthonormal transformation ψ as:

$$(G^{*T}\ \hat{C}_N\ G^*)^{-1}\ (G^{*T}\ \hat{C}_F\ G^*)\ \Psi = \Psi\ \Lambda_t\quad . \tag{A-3-90}$$

Using Eq. A-3-90 in Eq. A-3-89, we have the eigenvalue relationship:

$$\hat{C}_N^{-1}\ \hat{C}_F\ G^*\Psi = G^*\Psi\ \Lambda_t\quad .$$

Since any orthonormal transformation does not change the J-divergence, we can include Ψ into G^*. Therefore, optimum G^* should consist of the eigenvectors of $\hat{C}_N^{-1}\ \hat{C}_F$. Also, note that the eigenvalues of $(G^{*T}\ \hat{C}_N\ G^*)^{-1}\ (G^{*T}\ \hat{C}_F\ G^*)$ in the t-dimensional subspace are the same as the t eigenvalues of $\hat{C}_N^{-1}\ \hat{C}_F$ in the original subspace. Then $J_{FN}(p,t)$ becomes

$$J_{FN}(p,t) = \frac{1}{2}\sum_{r=1}^{t}\ (\lambda_r + \lambda_r^{-1} - 2)\quad .$$

Therefore, in order to maximize $J_{FN}(p,t)$, we choose G to be the t normalized eigenvectors of $\hat{C}_N^{-1}\ \hat{C}_F$ corresponding to the eigenvalues λ_r for which

$$\lambda_r + \lambda_r^{-1} \geq \lambda_q + \lambda_q^{-1}\ ;\quad q=1,2,\cdots,(p+1)m\ \ ;\quad r=1,2,\cdots,t\quad .$$

b. Maximization of Bhattacharyya Distance

The gradient expression for the Bhattacharyya distance of Eq. A-3-68, when $\bar{Y}_{pF}=0$ (to within a scale factor), is given by

$$\nabla_G B_{FN}(p,t) = [\hat{C}_N G (G^T \hat{C}_N G)^{-1} - \hat{C}_F G (G^T \hat{C}_F G)^{-1}]$$

$$\cdot [(G^T \hat{C}_N G)^{0.5}(G^T \hat{C}_F G)^{-0.5} - (G^T \hat{C}_N G)^{-0.5}(G^T \hat{C}_F G)^{0.5}] \quad . \tag{A-3-91}$$

Optimal G^* satisfies either

$$\hat{C}_N^{-1} \hat{C}_F G^* = G^* (G^{*T} \hat{C}_N G^*)^{-1}(G^{*T} \hat{C}_F G^*) \tag{A-3-92}$$

or

$$G^{*T} (\hat{C}_N - \hat{C}_F) G^* = 0 \quad . \tag{A-3-93}$$

The term G^* of Eq. A-3-93 cannot be optimum, because this makes $G^*=0$ and $B_{FN}=0$ (unless $\hat{C}_N=\hat{C}_F$, in which case it is indeterminate and a meaningless problem). Since Eq. A-3-92 is the same as Eq. A-3-89, G^* should consist of normalized eigenvectors of $\hat{C}_N^{-1} \hat{C}_F$. In terms of eigenvalues of $\hat{C}_N^{-1} \hat{C}_F$, the Bhattacharyya distance is given by

$$B_{FN}(p,t) = \frac{1}{2} \sum_{r=1}^{t} \ell n \left[\lambda_r^{0.5} + \lambda_r^{-0.5}/2\right] \quad .$$

Since $\ell n\ x$ is a monotonic function of x, $B_{FN}(p,t)$ is maximized if we select G^* to be the t eigenvectors of $\hat{C}_N^{-1} \hat{C}_F$ corresponding to the eigenvalues λ_r for which

$$\lambda_r^{0.5} + \lambda_r^{-0.5} > \lambda_q^{0.5} + \lambda_q^{-0.5} \quad ; \quad q=1,2,\cdots,(p+1)m \quad ; \quad r=1,2,\cdots,t \quad .$$

Noting that $\lambda_r + \lambda_r^{-1} = (\lambda_r^{0.5} + \lambda_r^{-0.5})^2 - 2$, we immediately see that the Bhattacharyya distance and the J-divergence yield identical parity transformations.

c. Maximization of I-Divergence

The gradient expression for the I-divergence of Eq. A-3-66 when $\bar{Y}_{pF} = 0$ is given by:

$$\nabla_G I_{FN}(p,t) = \hat{C}_N G (G^T \hat{C}_F G)^{-1} - \hat{C}_F G (G^T \hat{C}_F G)^{-1}(G^T \hat{C}_N G) \cdot$$

$$(G^T \hat{C}_F G)^{-1} + \hat{C}_F G (G^T \hat{C}_F G)^{-1}$$

$$- \hat{C}_N G (G^T \hat{C}_N G)^{-1}(G^T \hat{C}_F G)(G^T \hat{C}_N G)^{-1} . \qquad \text{(A-3-94)}$$

At the optimum G^*, we have

$$\hat{C}_N^{-1} \hat{C}_F G^* = G^* (G^{*T} \hat{C}_N G^*)^{-1}(G^{*T} \hat{C}_F G^*) . \qquad \text{(A-3-95)}$$

Using the same argument as before G^* is chosen from the eigenvectors of $\hat{C}_N^{-1} \hat{C}_F$. The value of the I-divergence in terms of the eigenvalues of $\hat{C}_N^{-1} \hat{C}_F$ is given by

$$I_{FN}(p,t) = \sum_{r=1}^{t} (\lambda_r - \ell n \lambda_r - 1) .$$

Since $\lambda_r > \ell n \lambda_r$ for all $\lambda_r > 0$, $I_{FN}(p,t)$ is maximized by selecting G to be the t normalized eigenvectors of $\hat{C}_N^{-1} \hat{C}_F$ corresponding to the t largest eigenvalues.

RELATIONSHIP WITH OTHER MEASURES

For the scale-factor failures, a variety of criteria similar to the I-divergence have been used in the literature. We will discuss four such criteria here. First in [13] a robust redundancy metric is used for detection, J_1, that involves maximization of the difference in the covariance of parity relations under failed and normal hypotheses:

$$J_1(p,t) = \text{tr}\ [G^T(\hat{C}_F - \hat{C}_N)\ G] \qquad \text{(A-3-96a)}$$

subject to

$$G^T G = I_t \quad . \qquad \text{(A-3-96b)}$$

Thus, if we interpret $\text{tr}(G^T\hat{C}_F G)$ as a measure of "failure signature" strength under failed hypothesis and $\text{tr}(G^T\hat{C}_N G)$ as a measure of noise strength, the criterion in [13] corresponds to maximizing the difference in energies between the failed and normal modes of system operation. In this case, the optimal parity transformation G is given by the t eigenvectors of $(\hat{C}_F - \hat{C}_N)$ corresponding to the largest t eigenvalues of $(\hat{C}_F - \hat{C}_N)$.

A second criterion is related to maximizing the difference in entropy of parity relations under failed and normal system operation:

$$J_2 = H_F(p,t) - H_N(p,t) = \frac{1}{2}\ \ell n\ \det\ (\hat{C}_N^{-1}\hat{C}_F) \quad . \qquad \text{(A-3-97)}$$

This criterion has precisely the same optimal parity transformation G^* as that obtained by maximizing the I-divergence. This criterion is used extensively, albeit in an ad hoc manner, in feature selection problems [22].

A third criterion is related to maximizing the signal strength subject to a constraint on the noise strength in the parity relations. The objective function is to maximize

$$J_3 = \text{tr}\ (G^T\hat{C}_F G) \qquad \text{(A-3-98a)}$$

subject to

$$\text{tr}\ (G^T\hat{C}_N G) = 1 \quad . \qquad \text{(A-3-98b)}$$

Here, $\mathrm{tr}(G^T\hat{C}_F G)$ is a measure of the detection capability of the parity relations and $\mathrm{tr}(G^T\hat{C}_N G)$ is a measure of the false alarm probability. It can be shown that the optimal parity transformation G is the same (except for a scale factor) as that obtained by maximizing the I-divergence.

Finally, continuing on our signal strength and noise strength interpretation of $\mathrm{tr}(G^T\hat{C}_F G)$ and $\mathrm{tr}(G^T\hat{C}_N G)$, respectively, a natural criterion is to choose G to maximize the signal to noise ratio:

$$J_4 = \frac{\mathrm{tr}\ (G^T\hat{C}_F G)}{\mathrm{tr}\ (G^T\hat{C}_N G)} . \qquad \text{(A-3-99)}$$

This is a very difficult measure to optimize. However, it is stated in [22] that the eigenvector solution obtained for the I-divergence criterion provides near-optimal solutions for this problem as well.

A FREQUENCY DOMAIN ALGORITHM

As is the case with the robust parity checks of Section A-2, the AR filter interpretation of parity relations and the asymptotic per sample distance measures of Eqs. A-3-35 and A-3-36 can be used to devise a frequency domain algorithm for the parity generation, whenever the normal and failure modes result in stationary models. We illustrate the frequency domain algorithm for the I-divergence measure. Similar algorithms are obtained for the J-divergence and the Bhattacharyya distance measures.

The asymptotic per sample I-divergence in the equal mean case, i.e., when $\bar{Y}_{pF}(k) = \bar{Y}_{pN}(k) = 0$, is obtained from Eq. A-3-35 as

$$\bar{I}_{FN} = (4\pi)^{-1} \int_0^{2\pi} \{tr[G^T(j\omega)\ \Phi_{yN}(\omega)\ G^*(j\omega)\big)^{-1}\big(G^T(j\omega)\ \Phi_{yF}(\omega)\ G^*(J\omega)\big)]$$

$$- \ell n \det\ [\big(G(j\omega)\ \Phi_{yN}(\omega)\ G^*(j\omega)\big)^{-1}\big(G^T(j\omega)\ \Phi_{yF}(\omega)\ G^*(j\omega)\big)]\} \quad \text{(A-3-100)}$$

An approximate algorithm for obtaining a p-th order parity relation is as follows:

a. Divide the interval $[0,2\pi]$ into p equal intervals and let

$$\omega_r = r\,\frac{2\pi}{p} \quad ; \quad r = 0,1,2,\cdots,p \quad .$$

b. As in Section A-2, evaluate $\Phi_{yN}(\omega)$ and $\Phi_{yF}(\omega)$ at ω_r and the t orthogonal eigenvectors, g_{ir}, $1 \leq i \leq t$ of dimension m corresponding to the t largest eigenvalues of $\Phi_{yN}^{-1}(\omega_r)\ \Phi_{yF}(\omega_r)$. Let

$$G_r = [g_{1r}, g_{2r}, \cdots, g_{tr}] \quad \text{an m by t matrix.} \quad \text{(A-3-101)}$$

c. Normalize each column of g_{ir} so that

$$\sum_{q=0}^{p} g_{ir}^T\ g_{ir} = 1$$

where g_{ir} is the i-th column of G_r. The resulting normalized g_{ir} vectors provide the (p+1)m by t parity transformation matrix G.

As in Section A-2, the result can be expanded to any order p with a linear growth in computation time.

A-3.3.3 General Failure Modes

The general case where a failure manifests as a bias and scale factor change requires a gradient scheme for optimization. However, suboptimum procedures can be generated following the approach proposed for feature selection problems [22].

MAXIMIZATION OF J-DIVERGENCE

In the general case, when means and covariances are unequal, we could use the following suboptimum procedure to maximize the J-divergence (Eq. A-3-67).

a. Form the eigenvectors of $\hat{C}_N^{-1}\,\hat{C}_F$. Let F be the set of eigenvectors and $d = F\,\bar{Y}_{pF}$. Let q-th component of d be d_q. It is well known that F can be selected such that it simultaneously diagonalizes $\hat{C}_N$ and $\hat{C}_F$ [22] as

$$F\,\hat{C}_N\,F^T = I \qquad F\,\hat{C}_F\,F^T = \Lambda \quad . \tag{A-3-102}$$

b. Choose the columns of parity transformation G to be those t columns of F for which

$$\left(1 + \lambda_r^{-1}\right) d_r^2 + \lambda_r + \lambda_r^{-1} > \left(1 + \lambda_q^{-1}\right) d_q^2 + \lambda_q + \lambda_q^{-1} \quad \text{for all}$$

$$q = 1,2,\cdots,(p+1)m \quad , \quad r = 1,2,\cdots,t \quad . \tag{A-3-103}$$

The divergence for t selected parity relations becomes

$$J_{FN}(p,t) = \sum_{r=1}^{t} \left\{\left(1 + \lambda_r^{-1}\right) d_r^2 + \lambda_r + \lambda_r^{-1} - 2\right\} \quad . \tag{A-3-104}$$

This approximation works well when the divergence term related to the mean can be expressed by a small number of t eigenvectors chosen according to Eq. A-3-104. Alternatively, when the mean related term is dominant, we can select

the first parity relation according to Eq. 3-70 with $\hat{C} = [\hat{C}_N + \hat{C}_F]/2$, and the remaining (t-1) parity relations from the eigenvectors of $\hat{C}_N^{-1}\hat{C}_F$ corresponding to the eigenvalues satisfying Eq. A-3-86. However, we lose the orthogonal property of the columns of G.

MAXIMIZATION OF THE BHATTACHARYYA DISTANCE

In the general case, when means and covarances of Y_p are unqual under the two hypotheses, a suboptimum procedure for choosing parity relations to maximize the Bhattacharyya distance (Eq. A-3-68) is as follows:

a. Form the eigenvector matrix F of $\hat{C}_N^{-1}\hat{C}_F$ such that

$$F\,\hat{C}_N\,F^T = I \quad ; \quad F\,\hat{C}_F\,F^T = \Lambda$$

Let

$$d = F\,\bar{Y}_{pF} \quad .$$

b. The columns of G are selected to be the columns of F for which

$$\frac{1}{2}(1+\lambda_r)^{-1}d_r^2 + \ell n\,(\lambda_r^{0.5} + \lambda_r^{-0.5}) > \frac{1}{2}(1+\lambda_q)^{-1}d_q + \ell n\,(\lambda_q^{0.5} + \lambda_q^{-0.5})$$

$$\text{for all } q = 1,2,\cdots,(p+1)m \quad ; \quad r = 1,2,\cdots,t \quad . \qquad \text{(A-3-105)}$$

The Bhattacharyya distance for the t selected parity relations becomes

$$B_{FN}(p,t) = \frac{1}{2}\sum_{r=1}^{t}\frac{1}{2}(1+\lambda_r)^{-1}\,d_r^2 + \ell n\,(\lambda_r^{0.5} + \lambda_r^{-0.5}/2) \quad . \qquad \text{(A-3-106)}$$

As with the divergence, this approximation works well when the term related to the mean can be expressed by a small number of t-eigenvectors chosen according to Eq. A-3-105. When the mean vector is dominant, we select the first parity

relation according to Eq. 3-70 with $\hat{C} = [\hat{C}_N + \hat{C}_F]/2$, and the remaining (t-1) parity relations as the eigenvectors of $\hat{C}_N^{-1} \hat{C}_F$ corresponding to the eigenvalues satisfying Eq. A-3-86. However, the columns of G are not orthogonal in the latter case.

MAXIMIZATION OF THE I-DIVERGENCE

In the general case, the maximization of the I-divergence in Eq. A-3-66 follows the procedure outlined below.

a. Form the eigenvector matrix such that

$$F \hat{C}_N F^T = I \quad ; \quad F \hat{C}_F F^T = \Lambda$$

Let

$$d = F Y_{pF} \quad .$$

b. Choose the columns of G as those columns of F for which

$$d_r^2 + \lambda_r - \ell n \lambda_r > d_q^2 + \lambda_q - \ell n \lambda_q \text{ for all}$$

$$q = 1,2,\cdots,(p+1)m \quad ; \quad r = 1,2,\cdots,t \quad . \qquad \text{(A-3-107)}$$

The I-divergence for the t selected parity relations is

$$I_{FN}(p,t) = \frac{1}{2} \sum_{r=1}^{t} \left(d_r^2 + \lambda_r + \ell n \lambda_r - 1\right) \quad . \qquad \text{(A-3-108)}$$

As with the divergence and Bhattacharyya distance this approximation works well when the term related to the mean can be expressed by a small number of eigenvectors chosen according to Eq. A-3-107. Alternatively, when the mean related term is dominant, we select the first parity relation according to Eq. A-3-70 with $\hat{C} = [\hat{C}_N + \hat{C}_F]/2$ and the remaining (t-1) parity relations as the eigenvectors of $\hat{C}_N^{-1} \hat{C}_F$ corresponding to the largest (t-1) eigenvalues. However, the columns of G are not orthogonal in the latter case.

It should be noted that the single failure results presented in this section are also applicable to multiple failure modes, if we treat all the failure modes besides the one we wish to isolate as uncertain normal-mode behavior.

A-3.4 EVALUATION OF ALTERNATIVE FDI SCHEMES

As discussed in Section A-1, a measure of effectiveness of an FDI is its ability to focus information among its residuals. That is, the effectiveness of an FDI system is related to its ability to "collect" relevant failure information quickly and accurately. Our information distance measures based on the I-divergence, the J-divergence, and the Bhattacharyya distance provide convenient tools to evaluate the effectiveness of alternative FDI schemes. In this section, we evaluate two different FDI schemes: one based on Kalman filter (KF) generated residuals and the other based on the parity space approach.

Consider a Kalman filter based on nominal system parameters (A_0, C_0, D_0, Q_0, R_0). Let the innovation representation of the system be given by

$$\hat{x}(k+1) = \bar{A}_{0k}\hat{x}(k) + K_f(k)\ \nu(k) \qquad \text{(A-3-109)}$$

$$\nu(k) = -C_0\hat{x}(k) + y(k) \qquad \text{(A-3-110)}$$

where

$$K_f(k) = A_0\Sigma_0(k|k-1)\ C_0^T\ (C_0\Sigma_0(k|k-1)\ C_0^T + R_0)^{-1} \qquad \text{(A-3-111)}$$

$$\Sigma_0(k+1|k) = (A_0 - K_f(k)\ C_0)\ \Sigma_0(k|k-1)\ (A_0 - K_f(k)\ C_0)^T$$

$$+ D_0 Q_0 D_0^T + K_f(k)\ R_0\ K_f^T(k) \qquad \text{(A-3-112)}$$

$$\bar{A}_{0k} = A_0 - K_f(k)\ C_0 \qquad \text{(A-3-113)}$$

$$\nu(k) = \text{Filter residuals} \qquad \text{(A-3-114)}$$

$$\hat{x}(k) = \text{Predicted state estimate} \quad . \qquad \text{(A-3-115)}$$

Suppose, we have an innovation sequence from the Kalman filter:

$$\nu^N = [\nu(1),\ \nu(2),\ \cdots,\ \nu(N)] \qquad \text{(A-3-116)}$$

where we have assumed, without loss of generality, that the sequence starts at time step k=1. Consider a similar sequence of parity relations generated from the parity space approach:

$$\mu^N = [\mu(1),\ \mu(2),\ \cdots,\ \mu(N)] \quad . \qquad \text{(A-3-117)}$$

Let the corresponding decision statistics (e.g., weighted sum-squared residuals, whitener-correlator, etc.) for detecting a failure F be $S_f(\nu^N)$ and $S_p(\mu^N)$, respectively. Then, the J-divergence and/or the Bhattacharyya distance of $S_f(\nu^N)$ between failed and unfailed modes of system operation, and similar distances for $S_p(\nu^N)$ provide measures for evaluating the two FDI schemes, in terms of their ability to distinguish the two hypotheses. We will illustrate the evaluation procedure for the whitener-correlator and the weighted sum-squared residual decision statistics. The evaluation procedure can be extended to multiple hypotheses in a straightforward manner.

A-3.4.1 Evaluation of Whitener-Correlator Decision Statistic

Assume that the Kalman filter based and the parity space based residuals be accumulated over time via a linear algorithm:

$$S_f(\nu^N) = \sum_{k=1}^{N} W_f^T(k)\ \nu(k) \qquad \text{(A-3-118a)}$$

$$S_p(\mu^N) = \sum_{k=1}^{N} W_p^T(k)\ \mu(k) \qquad \text{(A-3-118b)}$$

where $W_f(k)$ and $W_p(k)$ are selected to minimize some detection criterion. The problem is to evaluate the average distance measures over all uncertain models $\ell = 1,2,\cdots,L$ between the conditional distributions $p(S_f/F)$ and $p(S_f/N)$ for the Kalman filter based FDI scheme and $p(S_p/F)$ and $p(S_p/N)$ for the parity space based FDI scheme. These measures are given by

$$I_{FN}^{(i)} = \sum_{\ell=1}^{L} P_\ell\ I_{FN/\ell}^{(i)} \qquad i=f,p \qquad \text{(A-3-119)}$$

$$J_{FN}^{(i)} = \sum_{\ell=1}^{L} P_\ell\ J_{FN/\ell}^{(i)} \qquad i=f,p \qquad \text{(A-3-120)}$$

$$B_{FN}^{(i)} = \sum_{\ell=1}^{L} P_\ell\ B_{FN/\ell}^{(i)} \qquad i=f,p \qquad \text{(A-3-121)}$$

where the superscript i (=f,p) denotes that the evaluation is being made for filter based or parity based FDI schemes, respectively. The conditional distance measures based on model ℓ is:

$$I^{(i)}_{FN/\ell} = \frac{1}{2}\left[\ell n\left(\frac{\sigma^2_{Ni\ell}}{\sigma^2_{Fi\ell}}\right) + \left(\frac{\sigma^2_{Fi\ell}}{\sigma^2_{Ni\ell}}\right) - 1\right.$$

$$\left. + \left(\bar{S}_{Fi\ell} - \bar{S}_{Ni\ell}\right)^2 / \sigma^2_{Ni\ell}\right] \; ; \; i=f,p \qquad \text{(A-3-122)}$$

$$J^{(i)}_{FN/\ell} = \frac{1}{2}\left[\frac{\sigma^2_{Ni\ell}}{\sigma^2_{Fi\ell}} + \frac{\sigma^2_{Fi\ell}}{\sigma^2_{Ni\ell}} - 2\right.$$

$$\left. + \left(\bar{S}_{Fi\ell} - \bar{S}_{Ni\ell}\right)^2 \left(\frac{1}{\sigma^2_{Ni\ell}} + \frac{1}{\sigma^2_{Fi\ell}}\right)\right] \; ; \; i=f,p \qquad \text{(A-3-123)}$$

$$B^{(i)}_{FN/\ell} = \frac{1}{2}\left[\ell n\left\{\left(\frac{\sigma^2_{Fi\ell}}{\sigma^2_{Ni\ell}}\right)^{0.5} + \left(\frac{\sigma^2_{Ni\ell}}{\sigma^2_{Fi\ell}}\right)^{0.5}/2\right\}\right.$$

$$\left. + \frac{1}{4}\left(\bar{S}_{Fi\ell} - \bar{S}_{Ni\ell}\right)^2 / \left(\sigma^2_{Fi\ell} + \sigma^2_{Ni\ell}\right)\right] \; ; \; i=f,p \qquad \text{(A-3-124)}$$

where $\sigma^2_{Fi\ell}$ and $\sigma^2_{Ni\ell}$ are the variances of the decision statistic under failed and normal hypotheses, respectively for the filter based or parity based FDI scheme under model ℓ. Similarly, $\bar{S}_{Fi\ell}$ and $\bar{S}_{Ni\ell}$ are the corresponding mean values.

EVALUATION OF MEAN AND VARIANCE OF DECISION STATISTICS

The mean values of the decision statistics are given by

$$\bar{S}_{Ff\ell} = \sum_{k=1}^{N} W^T_f(k)\ \bar{\nu}_{F\ell}(k) \qquad \text{(A-3-125)}$$

where

$$\bar{\nu}_{F\ell}(k) = E_{\ell}\ \{\nu(k)/F\}\ . \qquad \text{(A-3-126)}$$

The mean $\bar{\nu}_{F\ell}(k)$ can be evaluated from the 2n-dimensional augmented system of equations (see Eqs. A-3-3, A-3-4, A-3-109).

$$\begin{bmatrix} x(k+1) \\ \hat{x}(k+1) \end{bmatrix} = \begin{bmatrix} A_{F\ell} & 0 \\ K_f(k)C_{F\ell} & \bar{A}_{0k} \end{bmatrix} \begin{bmatrix} x(k) \\ \hat{x}(k) \end{bmatrix} + \begin{bmatrix} D_{F\ell} \\ 0 \end{bmatrix} w(k) + \begin{bmatrix} I \\ 0 \end{bmatrix} d_F(k)$$

$$+ \begin{bmatrix} 0 \\ K_f(k) \end{bmatrix} \left[\nu_{F\ell}(k) + b_F(k)\right] \qquad \text{(A-3-127)}$$

$$\nu_{F\ell}(k) = \left[C_{F\ell} - C_0\right] \begin{bmatrix} x(k) \\ \hat{x}(k) \end{bmatrix} + \nu_{F\ell}(k) + b_F(k)\ . \qquad \text{(A-3-128)}$$

Denoting $\hat{x}_a^T(k) = \left(x(k)\ \hat{x}(k)\right)$, Eqs. A-3-127 and A-3-128 can be rewritten as

$$x_a(k+1) = A_{a\ell}(k)\ x_a(k) + D_{a\ell}\ w(k) + B_{a\ell}\ d_F(k)$$

$$+ K_{a\ell}(k)\ \left[\nu_{F\ell}(k) + b_F(k)\right] \qquad \text{(A-3-129)}$$

$$\nu_{F\ell}(k) = C_{a\ell}\ x_a(k) + \nu_{F\ell}(k) + b_F(k) \qquad \text{(A-3-130)}$$

where $A_{a\ell}$, $D_{a\ell}$, $B_{a\ell}$, $K_{a\ell}$ and $C_{a\ell}$ have obvious definitions from Eqs. A-3-127 and A-3-128. Mean $\bar{\nu}_{F\ell}(k)$ and its covariance $\Omega_{F\ell}(k)$ can be evaluated from Eqs. A-3-129 and A-3-130 as

$$\bar{\nu}_{F\ell}(k) = C_{a\ell}\, \bar{x}_a(k) + b_F(k) \tag{A-3-131a}$$

$$\Omega_{F\ell}(k) = C_{a\ell}\, \Sigma_{a\ell}(k)\, C_{a\ell}^T + R_{F\ell} \tag{A-3-131b}$$

where

$$\bar{x}_a(k+1) = A_{a\ell}(k)\, \bar{x}_a(k) + B_{a\ell}\, d_F(k) + K_{a\ell}(k)\, b_F(k) \tag{A-3-132a}$$

$$\Sigma_{a\ell}(k+1) = A_{a\ell}(k)\, \Sigma_{a\ell}(k)\, A_{a\ell}^T(k) + D_{a\ell}\, Q\, D_{a\ell}^T \tag{A-3-132b}$$

Similarly, one can compute $\bar{\nu}_{N\ell}(k)$ and $\Omega_{N\ell}(k)$ for the normal mode. The variance of the decision statistic $\sigma_{Ff\ell}^2$ is computed from

$$\sigma_{Ff\ell}^2 = \sum_{k=1}^{N} \sum_{j=1}^{N} W_f^T(k)\, E_\ell\{(\nu(k) - \bar{\nu}_{F\ell})\, (\nu(j) - \bar{\nu}_{F\ell}(j))^T/F\}\, W_f(j) \tag{A-3-133a}$$

$$= \sum_{k=1}^{N} \sum_{j=1}^{N} W_f^T(k)\, [E_\ell\{\nu(k)\, \nu^T(j)/F\} - \bar{\nu}_{F\ell}(k)\, \bar{\nu}_{F\ell}^T(j)]\, W_f(j)\ . \tag{A-3-133b}$$

When $\nu(k)$ sequence is uncorrelated, $\sigma_{Ff\ell}^2$ takes a simple form:

$$\sigma_{Ff\ell}^2 = \sum_{k=1}^{N} W_f^T(k)\, \Omega_{F\ell}(k)\, W_f(k)\ . \tag{A-3-134}$$

However, $\nu(k)$ is correlated in uncertain systems and, therefore, we must evaluate the cross-correlation terms of the form $E_\ell \left(\nu(k)\ \nu^T(j)/F\right)$. The computation of the cross-correlation term proceeds as follows. From Eqs. A-3-129 and A-3-130, we have

$$E_\ell \left\{\nu(k)\ \nu^T(j)/F\right\} = \begin{cases} C_{a\ell}\ \Phi_{a\ell}(k,j)\ \Sigma_a(j)\ C_{a\ell}^T & ;\ k > j \\ \Omega_{F\ell}(k) & ;\ k = j \\ C_{a\ell}\ \Sigma_a(k)\ \Phi_{a\ell}^T(j,k)\ C_{a\ell}^T & ;\ j > k \end{cases} \qquad \text{(A-3-135)}$$

where

$$\Phi_{a\ell}(k,j) = \prod_{r=j}^{k-1} A_{a\ell}(r) \quad .$$

Similar expressions can be obtained for the normal mode of operation by replacing subscripts $F\ell$ by $N\ell$ in Eqs. A-3-131 through A-3-135. Once $\left(\bar{S}_{Ff\ell},\ \sigma^2_{Ff\ell}\right)$ and $\left(\bar{S}_{Nf\ell},\ \sigma^2_{Nf\ell}\right)$ are obtained, it is a simple matter to compute the distance measures in Eqs. A-3-119 through A-3-124. It is straightforward to evaluate $\bar{S}_{Fp\ell}$ and $\sigma^2_{Fp\ell}$ for the parity space based decision statistic in Eq. A-3-118b. The result is

$$\bar{S}_{Fp\ell} = \sum_{k=1}^{N} W_p^T(k)\ G^T\ \bar{Y}_{pF\ell}(k) \qquad \text{(A-3-136)}$$

$$\sigma^2_{Fp\ell} = \sum_{k=1}^{N} \sum_{j=1}^{N} W_p^T(k)\ G^T \left[E_\ell \left\{Y_p(k)\ Y_p^T(j)|F\right\} - \bar{Y}_{pF\ell}(k)\ \bar{Y}^T_{pF\ell}(j)\right] G\ W_p(j) \quad . \qquad \text{(A-3-137)}$$

Although a little tedius, the mean term $\bar{Y}_{pF\ell}(k)$ and the correlation term can be computed from Eqs. A-3-3 and A-3-4 in a straightforward manner. Similar expressions hold for the normal mode of system operation.

A-3.4.2 Evaluation of Weighted Sum-Squared Residual (WSSR) Decision Statistic

The WSSR statistic for the two approaches is given by

$$S_f(\nu^N) = \sum_{k=1}^{N} \nu^T(k)\ W_f^{-1}(k)\ \nu(k) \qquad \text{(A-3-138)}$$

$$S_p(\mu^N) = \sum_{k=1}^{N} \mu^T(k)\ W_p^{-1}(k)\ \mu(k) \quad . \qquad \text{(A-3-139)}$$

For a given failure mode F and model ℓ, the statistics in Eqs. A-3-138 and A-3-139 have non-central chi-squared distribution. However, for the number of samples N > 15, the statistics in Eqs A-3-138 and A-3-139 are approximately Gaussian. The computation of mean values is relatively easy:

$$\bar{S}_{Ff\ell} = \sum_{k=1}^{N} \text{tr}\ [W_f^{-1}(k)\ (\Omega_{Ff\ell}(k) + \bar{\nu}_{Ff\ell}(k)\ \bar{\nu}_{Ff\ell}^T(k))] \qquad \text{(A-3-140)}$$

$$\bar{S}_{Fp\ell} = \sum_{k=1}^{N} \text{tr}\ [W_p^{-1}(k)\ G^T\ E_\ell\ \{Y_p(k)\ Y_p^T(k)/F\}G] \qquad \text{(A-3-141a)}$$

$$= \sum_{k=1}^{N} \text{tr}\ [W_p^{-1}(k)\ G^T\ (\hat{C}_F(k) + \bar{Y}_{pF}(k)\ \bar{Y}_{pF}^T(k))G] \quad . \qquad \text{(A-3-141b)}$$

The computation of covariance proceeds as follows:

$$\sigma^2_{Ff\ell} = E_\ell\ \{S_f^2\ (\nu^N/F)\} - \bar{S}^2_{Ff\ell} \qquad \text{(A-3-142a)}$$

$$\sigma^2_{Fp\ell} = E_\ell\ \{S_p^2\ (\mu^N/F)\} - \bar{S}^2_{Fp\ell} \quad . \qquad \text{(A-3-142b)}$$

We now compute the mean-squared value required in Eq. A-3-142:

$$E_\ell \{S_f^2(\nu^N)/F\} = E_\ell \{ \sum_{k=1}^{N} \sum_{j=1}^{N} \nu^T(k)\, W_f^{-1}(k)\, \nu(k)\, \nu^T(j)\, W^{-1}(j)\, \nu(j)/F \} \quad . \tag{A-3-143}$$

In order to evaluate the expectation in Eq. A-3-143, we need the result of Theorem 3.3.

Theorem 3.3 - Let x_1 and x_2 be Gaussian random vectors of dimension n with means $\bar{x}_1$ and $\bar{x}_2$. Their covariances are Σ_1 and Σ_2, and the cross-covariance between x_1 and x_2 is Σ_{12}. Then

$$\begin{aligned} E\{x_1^T A x_1\, x_2^T B x_2\} = {} & (\bar{x}_1^T A \bar{x}_1\, \bar{x}_2^T B \bar{x}_2) \\ & + 2\, tr\, (A \Sigma_{12} B \Sigma_{12}^T) \\ & + tr\, (A \Sigma_1)\, tr\, (B \Sigma_2) \quad . \end{aligned} \tag{A-3-144}$$

Proof: Note that

$$x_1^T A x_1\, x_2^T B x_2 = \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{\ell=1}^{n} \sum_{m=1}^{n} a_{ij}\, b_{\ell m}\, x_{1i}\, x_{1j}\, x_{2\ell}\, x_{2m} \quad . \tag{A-3-145}$$

In order to evaluate the expectation of Eq. A-3-145, we need to find the expectation of the product of four non-zero mean Gaussian random variables, $E\{x_{1i}\, x_{1j}\, x_{2\ell}\, x_{2m}\}$. To compute this expectation, we define

$$v^T = (v_1\ v_2\ v_3\ v_4) = (x_{1i}\ x_{1j}\ x_{2\ell}\ x_{2m})$$

$$\omega^T = (\omega_1\ \omega_2\ \omega_3\ \omega_4) \quad .$$

Let the covariance of v be Σ_v and is of the form

$$\Sigma_v = \begin{bmatrix} \sigma_{1ii} & \sigma_{1ij} & \sigma_{12i\ell} & \sigma_{12im} \\ \sigma_{1ij} & \sigma_{1jj} & \sigma_{12j\ell} & \sigma_{12jm} \\ \sigma_{12i\ell} & \sigma_{12j\ell} & \sigma_{2\ell\ell} & \sigma_{2\ell m} \\ \sigma_{12im} & \sigma_{12jm} & \sigma_{2\ell m} & \sigma_{2mm} \end{bmatrix}$$

The characteristic function of v is

$$\phi(\omega) = E\ \{\exp\ (-j\omega^T v)\}\ .$$

Using the fact that v is normal, we have

$$\Phi(\omega) = \exp\ \left(-j\omega^T \bar{v} - \frac{1}{2}\ \omega^T\ \Sigma_v\ \omega\right)\ .$$

From the property

$$\left.\frac{\partial^4\ \Phi(\omega)}{\partial\omega_1\ \partial\omega_2\ \partial\omega_3\ \partial\omega_4}\right|_{\omega=0} = j^4\ E\ \{v_1\ v_2\ v_3\ v_4\} = E\ \{x_{1i}\ x_{1j}\ x_{2\ell}\ x_{2m}\} \qquad \text{(A-3-146)}$$

we have

$$E\ \{x_{1i}\ x_{1j}\ x_{2\ell}\ x_{2m}\} = \bar{x}_{1i}\ \bar{x}_{1j}\ \bar{x}_{2\ell}\ \bar{x}_{2m}$$

$$+\ \sigma_{1ij}\ \sigma_{2\ell m} + \sigma_{12i\ell}\ \sigma_{12jm} + \sigma_{12im}\ \sigma_{12j\ell}\ . \qquad \text{(A-3-147)}$$

Using Eq. A-3-147 in Eq. A-3-145, we have

$$E\{x_1^T A x_1 \; x_2^T B x_2\} = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\ell=1}^{n}\sum_{m=1}^{n} a_{ij} b_{\ell m} (\bar{x}_{1i} \bar{x}_{1j} \bar{x}_{2\ell} \bar{x}_{2m}$$

$$+ \sigma_{1ij} \sigma_{2\ell m} + \sigma_{12i\ell} \sigma_{12jm} + \sigma_{12im} \sigma_{12j\ell})$$

$$= (\bar{x}_1^T A \bar{x}_1 \; \bar{x}_2^T B \bar{x}_2)$$

$$+ \text{tr}\,(A \Sigma_1 B \Sigma_2)$$

$$+ 2\,\text{tr}\,(A \Sigma_{12} B \Sigma_{12}^T) \quad .$$

Using the result of Theorem 3.3 in Eq. A-3-143, we have

$$E_\ell\{S_f^2(\nu^N)|F\} = \{\sum_k \text{tr}\,[W_f^{-1}(k) \bar{\nu}_{Ff\ell}(k) \bar{\nu}_{Ff\ell}^T(k)]\}^2$$

$$+ \{\sum_k \text{tr}\,[W_f^{-1}(k) \Omega_{Ff\ell}(k)]\}^2$$

$$+ 2 \sum_j \sum_k \text{tr}\,[W_f^{-1}(k) [E_\ell\{\nu(k) \nu^T(j)|F\} - \bar{\nu}_{Ff\ell}(k) \bar{\nu}_{Ff\ell}^T(j)]$$

$$W_f^{-1}(j) [E_\ell\{\nu(j) \nu^T(k)|F\} - \bar{\nu}_{Ff\ell}(j) \bar{\nu}_{Ff\ell}^T(k)]] \; . \qquad \text{(A-3-148)}$$

The required cross-correlation terms are defined in Eqs. A-3-129 through A-3-135. The computation of mean and covariance for the normal mode KF based FDI scheme and similar quantities for the parity space based FDI scheme is straightforward. Thus the statistical distance measures, introduced in this

section for parity optimization, provide convenient measures for the evaluation of alternative FDI approaches as well. Note, in particular, that the evaluation measures are nonstationary and, hence, we have a mechanism to determine how fast failure information is being accumulated in various FDI schemes.

REFERENCES

1. Evans, F.A., and J.C. Wilcox, "Experimental Strapdown Redundant Sensor Inertial Navigation System," Journal of Spacecraft and Rockets, Vol. 7, No. 9, pp. 1070-1074, Sept. 1970.

2. Gilmore, J.P., and R.A. McKern, "A Redundant Strapdown Inertial Reference Unit (SIRU)," Journal of Spacecraft and Rockets, Vol. 9, No. 9, pp. 39-47, Jan. 1972.

3. Willsky, A.S., and H.L. Jones, "A Generalized Likelihood Ratio Approach to the Detection and Estimation of Jumps in Linear Systems," IEEE Transactions on Automatic Control, AC-21, pp. 108-112, Feb. 1976.

4. Willsky, A.S., E.Y. Chow, S.B. Gershwin, C.S. Greene, P.K. Houpt, and A.L. Kurkjian, "Dynamic Model-Based Techniques for the Detection of Incidents on Freeways," IEEE Transactions on Automatic Control, No. 3, pp. 347-360, June 1980.

5. Gustafson, D.E., A.S. Willsky, and J.Y. Wang, "Final Report: Cardiac Arrythmia Detection and Classification Through Signal Analysis," The Charles Stark Draper Laboratory, Cambridge, Massachusetts, Report No. R-920, July 1975.

6. Willsky, A.S., J.J. Doyst, and B.S. Crawford, "Two Self-Test Methods Applied to an Inertial System Problem," Journal of Spacecraft and Rockets, Vol. 12, No. 7, July 1975, pp. 434-437.

7. Beard, R.V., "Failure Accommodation in Linear Systems Through Self Reorganization", Man Vehicle Lab, MIT, Cambridge, Massachusetts, Report MIT-71-1, February 1971.

8. Jones, H.L., "Failure Detection in Linear Systems," Ph.D. Thesis, Dept. of Aeronautics and Astronautics, MIT, Cambridge, Massachusetts, September 1973.

9. Messerole, J.S., Jr., "Detection Filters for Fault-Tolerant Control of Turbofan Engines," Ph.D. Thesis, MIT, Dept. of Aeronautics and Astronautics, June 1981.

10. Willsky, A.S., "A Survey of Design Methods for Failure Detection in Dynamic Systems," Automatica, Vol. 12, pp. 601-611, 1976.

11. Deckert, J.C., M.N. Desai, J.J. Deyst, and A.S. Willsky, "F-8 DFBW Sensor Failure Identification Using Analytic Redundancy," IEEE Transactions on Automatic Control, Vol. AC-22, No. 5, pp. 795-803, October 1977.

12. Chow, E.Y., and A.S. Willsky, "Issues in the Development of a General Design Algorithm for Reliable Failure Detection," Proc. IEEE Conference on Decision and Control, Albuquerque, NM, December 1980.

13. Lou, Xi-Chang, A.S. Willsky, and G.C. Verghese, "Failure Detection with Uncertain Models," American Control Conference, San Francisco, California, June 1983.

14. Potter, J.E., and Suman, M.C., "Thresholdless Redundancy Management with Arrays of Skewed Instruments," Integrity in Electronic Flight Control Systems, AGARDOGRAPH-224, pp. 15-1 to 15-25, 1977.

15. Desai, M. and A. Ray, "A Failure Detection and Isolation Methodology," 20th IEEE Conference on Decisiion and Control, San Diego, California, December 1981, pp. 1363-1369.

16. Beattie, E.C., et al, "Sensor Failure Detection System: Final Report," NASA CR-165515, PWA-5756-17, NASA-Lewis Research Center, Contract NA53-22481, August 1981.

17. Tou, J.T., and R.C. Gonzalez, Pattern Recognition Principles, Addison Wesley Publishers, Reading, Massachusetts, 1974.

18. Rao, C.R., "The Use and Interpretation of Principal Component Analysis in Applied Research," Sankhya, Vol. 26, pp. 329-358, 1964.

19. Moore, B.C., "Principal Component Analysis: Controllability, Observability, and Model Reduction," IEEE Transactions on Automatic Control, Vol. AC-16, No. 1, pp. 17-32, February 1981.

20. Priestly, M.B., T. Subba Rao, and H. Tong, "Applications of Principal Component Analysis and Factor Analysis in the Identification of Multivariable Systems," IEEE Transactions on Automatic Control, Vol. AC-19, No. 6, pp. 730-734, December 1974.

21. Hunt, B.R., "Digital Image Processing," Proc. IEEE, Vol. 63, pp. 593-708, April 1975.

22. Fukunaga, K., Introduction to Statistical Pattern Recognition, Academic Press, New York, 1972.

23. Kay, S.M., and S.L. Marple, Jr., "Spectrum Analysis - A Modern Perspective," Proc. IEEE, Vol. 69, No. 11, pp. 1380-1419, November 1981.

24. Cadzow, J.A., "Spectral Estimation: An Overdetermined Rational Model Equation Approach," Proc. IEEE, Vol. 70, No. 9, pp. 907-939, September 1982, (Special issue on Spectral Estimation).

25. Tufts, D.W., and R. Kumaresan, "Estimation of Frequencies of Multiple Sinusoides: Making Linear Prediction Perform Like Maximum Likelihood," Proc. IEEE, Vol. 70, No. 9, pp. 975-989, September 1982, (Special issue on Spectral Estimation).

26. Beattie, E.C., et al, "Sensor Failure Detection for Jet Engines," Final Report, NASA CR-168190, May 1983.

27. Max, M., and T. Kailath, "Efficient Inversion of Toeplitz-Block Toplitz Matrix," IEEE Transactions of ASSP, Vol. 31, No. 5, pp. 1218-1231, October 1983.

28. Ohnishi, K., "Direct Recursive Estimation of Noise Statistics," in Control and Dynamic Systems: Advances in Theory and Application, C.T. Leondes, ed., Academic Press, New York, 1980.

29. Hawkes, R.M., and J.B. Moore, "Performance Bounds for Adaptive Estimation," Proc. IEEE, Vol. 64, No. 8, pp. 1143-1150, August 1976.

30. Anderson, B.D.O., J.B. Moore, and R.M. Hawkes, "Model Approximations via Prediction Error Identification," Automatica, Vol. 14, No. 6, pp. 615-622, November 1978.

31. Baram, Y., and N. Sandell, Jr., "Consistent Estimation of Finite Parameter Sets with Application to Linear System Identification," IEEE Transactions on Automatic Control, Vol. AC-23, pp. 451-454, June 1978.

32. Kailath, T., "The Divergence and Bhattachanyya Distance Measures in Signal Selection," IEEE Transactions on Communication Technology, Vol. COM-15, No. 1, pp. 52-60, February 1967.

33. Kazakos, D., and P. Papantoni-Kazakos, "Spectral Distance Measures Between Gaussian Processes," IEEE Transactions on Automatic Control, Vol. 25, No. 5, pp. 950-959, October 1979.

34. Kazakos, D., "On Resolution and Exponential Discrimination Between Gaussian Stationary Vector Processes and Dynamic Models," IEEE Transactions on Automatic Control, Vol. 25, No. 2, pp. 294-296, April 1980.

35. Tildesley, "A First Study of the Burmese Skull," Biometrica, Vol. 13, pp. 176-262, 1921.

36. Makusita, K., "On the Notion of Affinity of Several Distributions and Some of its Applications," American Inst. Statistic Math, Vol. 19, pp 181-192, 1967.

37. Shumway, R.H., and A.N. Unger, "Linear Discriminant Functions for Stationary Time-Series," Journal of American Statist. Ass., Vol. 69, pp. 948-956, December 1974.

38. Bhattacharyya, A., "On a Measure of Divergence Between Two Statistical Polulations Defined by Their Probability Distributions," Bull. Calcutta, Math. Soc., Vol. 35, pp. 99-109, 1943.

39. Kullback, S., Information Theory and Statistics, Dover Pub., New York, 1959.

40. Jeffreys, H., "An Invariant Form for the Prior Probability in Estimation Problems," Proc. Royal Statistical Soc. A., Vol. 186, pp. 453-461, 1946.

41. Jeffreys, H., Theory of Probability, Oxford University Press, 1948.

42. Mehra, R.K., "Optimal Input Signals for Parameter Estimation in Dynamic Systems - Survey and New Results," IEEE Transactions on Automatic Control, Vol. 19, pp. 753-768, December 1974.

43. Kleinman, D.L., and K.R. Pattipati, "An Information Matrix Approach for Aircraft Parameter Insensitive Control," Proc. 1977 Joint American Automatic Control Conference, New Orleans, LA, 1977.

44. Lainiotis, D.G., "A Class of Upper Bounds on Probability of Error for Multihypotheses Pattern Recognition," IEEE Transactions on Information Theory, Vol. 15, pp. 730-731, 1969.

APPENDIX B

ROBUST ACCOMMODATION

APPENDIX B

ROBUST ACCOMMODATION

Our general methodology and specific analytical results have focussed on robust failure detection and isolation (FDI) as the most important aspect of the overall FDIA problem. In this appendix, many of these results are interpreted and extended to solve the robust accommodation problem.

The sensor accommodation problem is one which ultimately must include the performance goals and trade-offs of the entire system under consideration. Excellent examples of these kind of considerations are given in [3] and [11]. One appealing means of accommodating sensor failures involves the estimation of the variables which are no longer available. This method is employed in [3] and [11], however control system modifications are also required because of the inaccuracies of the particular estimation algorithm which was employed.

As in the case of FDI, modeling errors play a large role in the ultimate performance of any estimation algorithm. Therefore, estimators employed in accommodation algorithms should be designed for insensitivity to modelling errors where possible.

The first step in designing any estimation algorithm is the determination of a model for the desired variables.

Consider the discrete time state space description,

$$x(k+1) = A\,x(k) \tag{B-1}$$

$$y(k) = C\ x(k) + n(k) \tag{B-2}$$

Input terms in B-1 and B-2 can be included in the results which follow and similar results derived.

In Section 4 and Appendix A, we discussed the formation of parity checks based on such models that take the form $\nu(k) = GY_p(k)$, where $Y_p(k)$ is a p-window of m dimensional observations, y(k), and G is the tx(p+1)m parity check matrix. Furthermore, we showed how each parity check (which is normally a zero mean white noise sequence) could be interpreted as either an ARMA model of the ith element of y(k), y_i, or as a state space model of a p-window of $y_i(k)$.

When modeling errors exist in the parameters (A,C) of Eq. B-1, the robust parity checks can be interpreted as robust ARMA or state space models for $y_i(k)$. These can then be used in an estimation algorithm as described below.

In order to estimate sensor values both during no-failure and with a bad sensor, we can design an estimator which is; 1) robust to modeling errors by using parity checks as ARMA models, and 2) uses only those sensors which are validated by the FDI algorithm. Let the parity check which is used for estimating y_i be written in terms of the y_i ARMA model;

$$y_i(k) = \sum_{j=I}^{p} \omega_I\ y_i(k-j) + \sum_{\ell \neq i} \sum_{j=0}^{p} \mu_j\ y_\ell(k-j) \tag{B-3}$$

Then, during normal operation, we have a closed loop estimator for all sensor values y_i, i=1,...,N-sensors;

$$\hat{y}_i(k) = \bar{y}_i(k) + K[y_i(k) - \bar{y}_i(k)] \tag{B-4}$$

where

$$\bar{y}_i(k) = \sum_{j=1}^{p} \omega_j \hat{y}_i(k-j) + \sum_{\ell \neq i} \sum_{j=0}^{p} \mu_j y_\ell(k-j) \tag{B-5}$$

When sensor y_i is identified as failed we must change the estimates for all sensor values. For y_i, we perform the "open loop" scheme,

$$\hat{y}_i(k) = \sum_{j=1}^{p} \omega_j \hat{y}_i(k-j) + \sum_{\ell \neq i}^{p} \sum_{j=0}^{p} \mu_j y_\ell(k-j) \tag{B-6}$$

For y_ℓ such that $i \neq \ell$ we use ARMA models (parity checks) for y_ℓ that don't include y_i in a closed loop estimator of the form (2) and (3), or, use the open loop estimate for y_i in computing $\hat{y}_\ell$.

Although the interpretation of parity checks as robust ARMA models can be utilized in forming models for various estimation algorithms, a more direct solution to the problem of finding robust ARMA models for the measured variables can be developed. Consider the state space description

$$x(k+1) = A\, x(k) \tag{B-7}$$

$$y(k) = C\, x(k) \tag{B-8}$$

with $p(A = A_\ell, B = B_\ell) = p_i$, and suppose we wish to find a pth order AR model for the vector y(k);

$$\hat{y}(k) = \sum_{j=1}^{p} \Phi_j\, y(k-j) \tag{B-9}$$

where Φ_j is an mxm matrix.

As in our approach to parity generation, we would like a metric which is related to estimation accuracy and is easily optimized by choice of the Φ_j. Consider the mean square estimation error defined by

$$J = E_\ell ||y(k) - \hat{y}(k)||^2 \tag{B-10}$$

where the $||\cdot||$ operator is the standard Euclidean distance.

Rewriting Eq. B-10 in terms of the state variable x(k-p) using B-7 and B-8, we have,

$$J = E_\ell \{ ||[C_\ell A_\ell^p - \sum_{j=1}^{p} \Phi_j C_\ell A_\ell^{p-j}] x||^2 \} \tag{B-11}$$

Since we are interested in solutions which are independent of the state, x, the problem of choosing Φ_j reduces to,

$$\min_{\Phi_j} \; E_\ell \{ ||C_\ell A_\ell^p - \sum_{j=1}^{p} \Phi_j \, C_\ell A_\ell^{p-j}||^2 \} \tag{B-12}$$

The expectation operator can be taken inside, the gradient with respect to Φ_j taken and set equal to zero yielding the following.

$$\text{Let,} \qquad M_\ell \triangleq \begin{bmatrix} C_\ell A_\ell^{p-1} \\ C_\ell A_\ell^{p-2} \\ \vdots \\ C_\ell \end{bmatrix} \tag{B-13}$$

$$b_i{}^\ell \triangleq \text{ith row of } C_\ell A_\ell^p \tag{B-14}$$

C 3

APPENDIX C

SOFTWARE DESCRIPTION

$$\Phi \triangleq [\Phi_1 | \Phi_2 | \cdots | \Phi_p] \triangleq \begin{bmatrix} ---\phi_1--- \\ ---\phi_2--- \\ ---\phi_m--- \end{bmatrix} \tag{B-15}$$

Define,

$$\bar{C} = E_\ell \{M_\ell M_\ell^T\} \tag{B-16}$$

$$\bar{b}_i = E_\ell \{M_\ell b_i^\ell\} \tag{B-17}$$

Then the solution which minimizes D-12 is given by the equation

$$\bar{C}\ \phi_i = \bar{b}_i \tag{B-18}$$

which is easily solved for ϕ_i providing $\bar{C}$ is nonsingular. When no modeling errors are present, $\bar{C}$ is less then full rank implying that more than one solution to D-18 is possible.

Finally we note that the above result is strongly related to the Levinson filtering algorithm [17] which utilizes the assumption of stationarity to derive a statistically based metric that results in linear equations for the AR coefficients and are based on the Toeplitz autocovariance matrix of the sequence y(k).

APPENDIX C

SOFTWARE DESCRIPTION

The robust FDI software is designed to be used for the development and evaluation of detection algorithms for sensor failures in dynamic systems. It was developed on ALPHATECH's VAX 11-750, in VAX-11 FORTRAN-77.

The software consists of four separately executable programs: 1) DESIGN computes parity check vectors, by one of several methods, from the dynamic system matrices; 2) PARCEV evaluates the parity check vectors for their ability to distinguish sensor failures from each other and from no failure; 3) KFRES evaluates the ability of a Kalman filter to distinguish sensor failures from each other and from no failure; and 4) SIMULATE simulates the states, controls and outputs (sensor measurements) of the dynamic system over time, including sensor failures if desired, and computes residuals, using parity vectors or a Kalman filter, in order to detect and identify sensor failures. The software is modular and well-commented for easy modification, checks for faulty input data that would otherwise cause a program to crash, and alerts the user to problems encountered during execution, such as an attempt to invert a singular matrix. Each program is now discussed in detail.

C-1. PROGRAM DESIGN

The dynamic system equations underlying all of the robust FDI software are of the form:

$$\underline{x}(k+1) = A\,\underline{x}(k) + B\,\underline{u}(k) \qquad \text{(C-1a)}$$

$$\underline{y}(k) = C\,\underline{x}(k) + D\,\underline{u}(k) + \underline{v}(k) \qquad \text{(C-1b)}$$

$$u_i(k) = \sqrt{q_i} \cdot N(0, 1) \qquad \text{(C-1c)}$$

with white noise vectors:

$$v_i(k) = \sqrt{r_i} \cdot N(0,1) \qquad \text{(C-1d)}$$

where $\underline{x}$ is the state vector, $\underline{u}$ is the vector of controls, $\underline{y}$ is the output vector; A, B, C and D are the system matrices; $\underline{r}$ is the measurement noise covariance vector, and $\underline{q}$ is the control noise covariance vector. In order to make use of these equations, several parameters must be defined. First, the dimensions of the system are needed: the number of states NS, the number of controls NC, and the number of outputs NO; these parameters determine the sizes of the system matrices and noise vectors, as well as the state, control and output vectors. Also, the order NP of the desired parity check vectors, which determines the time window length to be used in residual generation, must be specified. Finally, the number of system models NL is needed; this refers to the fact that randomly perturbed system matrices, A_ℓ, B_ℓ, C_ℓ, and D_ℓ are computed from the true matrices for $\ell = 1, \cdots, NL$, where, for example

$$A_\ell = A + \delta A \qquad \text{(C-2a)}$$

and δA is a matrix of zero-mean Gaussian perturbations. Each system model has an associated probability P_ℓ of being true, such that:

$$\sum_{\ell=1}^{NL} P_\ell = 1 \qquad \text{(C-2b)}$$

In program DESIGN, the formation of parity vectors from the system matrices and noise vectors proceeds in two steps: first, a composite system matrix C_o is formed; and second, the parity vectors are computed from the C_o matrix. Each of these two steps may be performed by two different methods. The C_o matrix may be formed by either the Null Space method or the Auto-Correlation Function method; while the parity vectors may be computed by either the Robust Residual method or the Robust Detection method. Each of these methods is described below.

Both methods of computing the C_o matrix begin by forming two matrices M_ℓ and Γ_ℓ from randomly perturbed system matrices, as follows:

$$M_\ell = \begin{bmatrix} C_\ell \\ C_\ell A_\ell \\ C_\ell A_\ell^2 \\ \vdots \\ C_\ell A_\ell^{NP} \end{bmatrix} \qquad \text{(C-3a)}$$

$$\Gamma_\ell = \begin{bmatrix} D_\ell & 0 & 0 & \cdots & 0 \\ C_\ell B_\ell & D_\ell & 0 & \cdots & 0 \\ C_\ell A_\ell B_\ell & C_\ell B_\ell & D_\ell & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ C_\ell A_\ell^{NP-1} B_\ell & C_\ell A_\ell^{NP-2} B_\ell & C_\ell A_\ell^{NP-3} B_\ell & \cdots & D_\ell \end{bmatrix} \qquad \text{(C-3b)}$$

If the Null Space method is being used, the matrix N_ℓ is then formed:

$$N_\ell = \begin{bmatrix} M_\ell & \Gamma_\ell \\ 0 & I \end{bmatrix} \qquad \text{(C-4a)}$$

and C_o is computed by the following sum:

$$C_o = \sum_{\ell=1}^{NL} P_\ell \cdot N_\ell N_\ell^T \quad . \qquad \text{(C-4b)}$$

If, however, the Auto-Correlation Function method is used, then C_o is set equal to the Auto-Correlation Function matrix (ACF), which is computed by a more complicated procedure. First, the output and control noise covariance matrices are set:

$$R_p = \begin{bmatrix} R & 0 & 0 & \cdots & 0 \\ 0 & R & 0 & \cdots & 0 \\ 0 & 0 & R & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & R \end{bmatrix} \qquad \text{(C-5a)}$$

$$C_u = \begin{bmatrix} Q & 0 & 0 & \cdots & 0 \\ 0 & Q & 0 & \cdots & 0 \\ 0 & 0 & Q & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & & \cdots & Q \end{bmatrix} \qquad \text{(C-5b)}$$

where R_p contains NP+1 copies of R, Cu contains NP+1 copies of Q, and

$$Q = \text{diag}\ [\ \underline{q}\ ] \tag{C-5c}$$

$$R = \text{diag}\ [\ \underline{r}\ ]\ . \tag{C-5d}$$

Next, for each system model ℓ, the following discrete time Lyapunov equation must be solved for C_{x_ℓ}, the state covariance matrix:

$$C_{x_\ell} = A_\ell C_{x_\ell} A_\ell^T + B_\ell Q B_\ell^T\ . \tag{C-6}$$

Then, the output covariance matrix (C_{y_ℓ}) and the cross-covariance matrix (C_{yu_ℓ}) can be computed:

$$C_{y_\ell} = M_\ell C_{x_\ell} M_\ell^T + \Gamma_\ell C_u \Gamma_\ell^T + R_p \tag{C-7a}$$

$$C_{yu_\ell} = \Gamma_\ell C_u\ . \tag{C-7b}$$

Finally, the Auto-Correlation Function matrix can be formed:

$$ACF = \sum_{\ell=1}^{NL} P_\ell \cdot \begin{bmatrix} C_{y_\ell} & C_{yu_\ell} \\ C_{yu_\ell}^T & C_u \end{bmatrix} \tag{C-8}$$

and the C_o matrix may be set equal to the ACF matrix.

Once the C_o matrix has been formed, using either the Null Space or the ACF method, the parity vectors may be computed in two ways. Under the Robust Residual method, the eigenvalues (λ_j) and normalized eigenvectors ($\underline{w}_j$) of C_o are computed, and the eigenvectors with smallest eigenvalues are used as

parity check vectors. If the Robust Detection method is used, then for j = 1,•••,NO the vector $\underline{b}_j$ must be formed as

$$\underline{b}_j = \begin{bmatrix} \underline{e}_j \\ \underline{e}_j \\ \vdots \\ \underline{e}_j \\ \underline{0} \\ \underline{0} \\ \vdots \\ \underline{0} \end{bmatrix} \qquad \text{(C-9)}$$

where $\underline{b}_j$ contains NP+1 copies of the unit vector $\underline{e}_j$ in the j-th direction of size NO, and NP+1 copies of the zero vector of size NC. Then, the parity vectors and associated values are computed as follows:

$$\underline{w}_j = \frac{C_o^{-1}\ \underline{b}_j}{\| C_o^{-1}\ \underline{b}_j \|} \qquad \text{(C-10a)}$$

$$\lambda_j = \underline{w}_j^T\ C_o\ \underline{w}_j \quad . \qquad \text{(C-10b)}$$

Note that both $\underline{b}_j$ and $\underline{w}_j$ have dimension (NP+1)(NO+NC) (like the dimensions of the C_o matrix). In both methods, small values of λ_j indicate that residuals computed from $\underline{w}_j$ will tend to be close to zero under normal operating conditions.

Another desirable feature of parity vectors is that they have a reasonably large "signal-noise ratio." The signal-noise ratio of a parity vector $\underline{w}_i$ is computed by DESIGN as:

$$S/N_{ij} = \frac{\underline{w}_i^T \underline{b}_j \cdot BFM_j}{\sqrt{\underline{w}_i^T (ACF) \underline{w}_i}} \qquad (C\text{-}11)$$

for each sensor output $j = 1, \cdots, NO$, where BFM_j is the bias failure magnitude for the j-th sensor.

The basic structure of subroutine calls by DESIGN is given in Fig. C-1, although no attempt has been made to show multiple calls to the same routine, and calls to standard library routines for matrix operations are not shown. Table C-1 gives a brief description of the purpose of each subroutine and function (including library routines). The inputs and outputs to DESIGN are described below.

A sample input file to DESIGN is shown in Fig. C-2. The first line of data read in is for NS, NO and NC, the number of states, outputs and controls. Next to be read in are NPMAX the maximum order of the parity checks, and NP_ALL, a flag indicating whether to use all values of NP; if NP_ALL is true, DESIGN computes parity checks of order NP = 0, ···,NPMAX. Then N_PAR_CHECK specifies the number of parity checks desired; if the Robust Residual method is used, these will be the ones with lowest eigenvalues; if the Robust Detection method is used, they will be the ones of lowest order. (Usually, if the Robust Detection method is used, N_PAR_CHECK should be set to (NPMAX+1)NO in order to print all of the parity checks generated.)

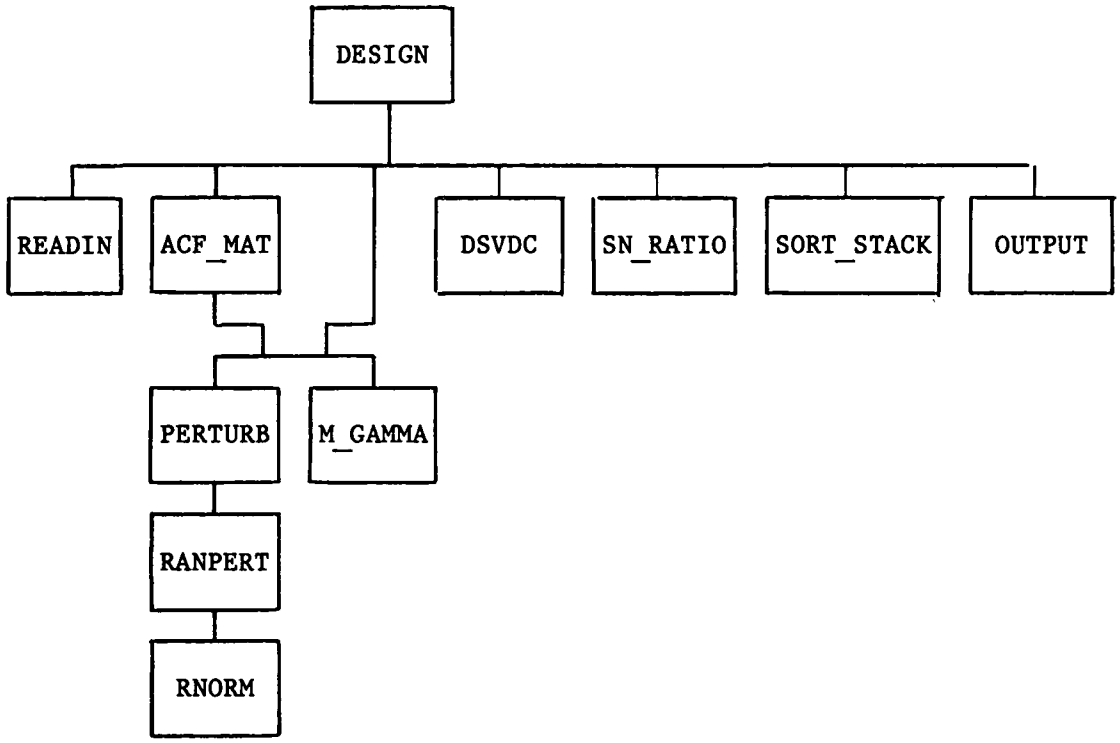


Figure C-1. Subroutine Calls Made by Program DESIGN, Excluding Library Routines

The next two lines of input data specify the generation of perturbed system matrices and the methods of parity check generation to be used. First, NL, TRUE_MODEL and SEEDO are read in. NL is the number of randomly perturbed system models used (the models are assumed to have equal probability of being correct, so $P_\ell = 1/NL$, for $\ell = 1, \cdots, NL$). TRUE_MODEL is a logical variable (true or false) indicating whether the first system model (ℓ=1) should use the unperturbed, correct system matrices; if TRUE_MODEL is false, the matrices for the first system model are randomly perturbed, as in the other models. SEEDO

TABLE C-1a. SUBROUTINES CALLED

Below is a brief description of the purpose of every subroutine called by DESIGN, PARCEV, KFRES or SIMULATE. See Figs. 1 3, 4 and 6 for the call structures.

ACF_MAT:	computes the Auto-Correlation Function (ACF) matrix (see Eq. E-8)
DSVDC:	LINPACK library singular value decomposition routine (used to compute eigenvalues and eigenvectors of symmetric C_0 matrix)
M_GAMMA:	computes the matrices M and Γ (see Eq. E-3)
OUTPUT:	writes the output files for DESIGN, including parity vectors and signal-noise ratios
PERTURB:	creates a perturbed version of the system matrices
PLOT:	ALPHATECH PLOT library routine that creates data files for later plotting
PROB-DIST:	computes the Bhattacharyya distance and correlation between two multivariate Gaussian probability distributions (see Eq. E-15)
RANPERT:	function that computes a random perturbation
READIN:	reads the DESIGN inputs from an input file and echoes to an output file
READ_KFR:	reads the KFRES input file
READ_PAR:	reads the PARCEV input file
READ_SIM:	reads the SIMULATE input file
RNORM:	returns a random number from a Gaussian normal distribution
SN_RATIO:	computes the signal-noise ratios for the most recently computed parity vector and each sensor output
SORT_STACK:	sorts an array of real numbers, keeping an array of integers in parallel
SPRT:	computes sequential probability ratio statistics (see Appendix B)
SYS_COVAR:	computes the covariance of the Kalman filter innovations for a given (perturbed) system model (see Eq. E-21b)
WSSR:	computes weighted sum of squared residuals statistics (see Appendix B)

TABLE C-1b. LIBRARY ROUTINES

The following are library routines for matrix manipulation used by DESIGN, PARCEV, KFRES and SIMULATE (not shown in Figs. C-1, 3, 4, and 6).

DGECOM:	LQGALPHA library routine used to compute the determinant of a matrix
MADD:	LQGALPHA library routine for adding two matrices
MLINEQ:	LQGALPHA library routine for solving linear equations (used for matrix inversion)
MMUL:	LQGALPHA library routine for multiplying two matrices
MSUB:	LQGALPHA library routine for subtracting one matrix from another
RICDSD:	LQGALPHA library routine for solving Riccati equations
SAVE:	LQGALPHA routine for copying a matrix into a part of another matrix
TRNATB:	LQGALPHA library routine for computing the transpose of a matrix

is the initial integer value of the random number generator seed. Next, the logical variables ACF_FLAG and RD_FLAG are read in. ACF_FLAG indicates whether to set the C_o matrix equal to the Auto-Correlation Function (ACF) matrix; if ACF_FLAG is false, the Null Space method of computing C_o is used. (The ACF matrix is computed in either case since it is needed to compute the signal-noise ratio.) RD_FLAG indicates whether the parity vectors should be computed by the Robust Detection method; if RD_FLAG is false, the Robust Residual method of computing parity vectors is used.

```
----------------------   'DESIGN' INPUTS   ----------------------------
   ---   NS, NO, NC      ---
    4    5    5
   ---   NPMAX, NP_ALL   ---
    0    T
   ---   N_PAR_CHECK     ---
    5
   ---   NL, TRUE_MODEL, SEED0    ---
    20   F    239457
   ---   ACF_FLAG, RD_FLAG    ---
    F    T
   ---   Q(NC)     ---    (346 RPM, 0, 0, 0, 0)
    5.343E-04   1.0E-08   1.0E-08   1.0E-08   1.0E-08
   ---   R(NO)     ---    (14 RPM, 12 RPM, .67 PSI, .09PSI, 2 DEG F)
    1.96E-06    6.49E-07  1.5E-06   4.86E-07  1.82E-06
   ---   RPM(NO)   ---
    5.0000000E-03  3.3333333E-03  5.4545455E-03  2.3076923E-02  5.0000000E-03
   ---   DEL_T     ---
    0.02
   ---   A (NS,NS)    ---
    0.8776392      9.5448047E-02 -7.6181483E-03  3.2125510E-02
    2.3492503E-03  0.9334235     2.9041661E-02 -8.4782876E-02
    3.8686348E-04 -1.4044064E-04  0.9854547     1.1503949E-04
    3.9211614E-03 -9.3381410E-04 -7.8085437E-04  0.9563294
   ---   B (NS,NC)    ---
    4.4787161E-02  6.4114124E-02 -1.5856765E-02  2.9632282E-03 -3.6747955E-02
    2.1774249E-02  1.4993310E-02 -1.2106993E-03 -2.1525344E-03 -1.2677516E-02
    1.4233845E-03  9.6010260E-04 -7.8124125E-05 -2.6199684E-06  9.1238425E-04
    1.7201110E-03  5.1628816E-04  2.1082116E-04 -1.6600619E-05  3.7244798E-03
   ---   C(NO,NS)    ---
    1.0000000      0.0000000E+00  0.0000000E+00  0.0000000E+00
    0.0000000E+00   1.000000      0.0000000E+00  0.0000000E+00
   -3.6199179E-02  0.8662740     -1.7536881E-02 -1.4256181E-02
    0.2380691      4.1814116E-03 -2.3846535E-02 -2.1576596E-02
   -1.380629      -0.6447408     -0.2106639      2.4134478E-02
   ---   D(NO,NC)    ---
    0.0000000E+00  0.0000000E+00  0.0000000E+00  0.0000000E+00  0.0000000E+00
    0.0000000E+00  0.0000000E+00  0.0000000E+00  0.0000000E+00  0.0000000E+00
    0.2299521     -0.4200383     2.9931962E-02  6.6308138E-03 -0.7154748
    0.1144538     -0.5347052     4.3202233E-02  1.8882763E-04  1.212762
    0.9383181      0.5814369    -1.8835608E-02 -3.4454153E-03  0.3496292
   ---   DEL_A    ---
    1.5380000E-02  2.1089999E-02  1.4011107E-03  1.2169647E-03
    3.1333333E-03  9.1000004E-03  8.7536051E-04  1.9486381E-04
    1.7636009E-03  1.1241082E-03  2.4200000E-03  1.1208000E-03
    4.4277371E-03  2.8102705E-03  2.7037999E-03  7.2799991E-03
   ---   DEL_B    ---
    8.5195694E-03  1.6469174E-03  5.1843708E-04  2.2065216E-04  8.6416223E-03
    2.4267395E-03  4.5617996E-04  1.5556102E-04  8.5836473E-05  2.6413030E-03
    8.5665914E-04  1.7833420E-04  5.5528384E-05  1.6012165E-05  8.5322355E-04
    2.1998493E-03  4.4917196E-04  1.3978184E-04  3.9906063E-05  2.1532946E-03
   ---   DEL_C    ---
    0.0000000E+00  0.0000000E+00  0.0000000E+00  0.0000000E+00
    0.0000000E+00  0.0000000E+00  0.0000000E+00  0.0000000E+00
    6.8665206E-02  0.1231549     3.7269455E-03  4.6490412E-03
    3.6485530E-02   1.930634     1.9171619E-03  3.1758063E-03
    0.1172195      9.3263514E-02  9.9560004E-03  1.3030000E-02
   ---   DEL_D    ---
    0.0000000E+00  0.0000000E+00  0.0000000E+00  0.0000000E+00  0.0000000E+00
    0.0000000E+00  0.0000000E+00  0.0000000E+00  0.0000000E+00  0.0000000E+00
    1.1518222E-02  3.2224100E-02  4.0293591E-03  3.9286349E-03  0.1696330
    6.5602068E-03  2.4839702E-02  3.7279809E-03  2.7165335E-04  4.1060939E-02
    6.2302982E-02  4.0404614E-02  8.7860543E-03  3.3902435E-03  0.2848548
```

Figure C-2. Inputs to Program DESIGN.

The next two inputs are the noise parameters. Q is an NC-vector which is the diagonal of the control noise covariance matrix; R is an NO-vector which is the diagonal of the measurement noise covariance matrix. (Off-diagonal elements of both matrices are zero.) BFM is an NO-vector which contains the bias failure magnitude for each sensor; and DEL_T is the time increment.

The final inputs to DESIGN are the correct system matrices, A, B, C and D, and the matrices containing the standard deviations of each element for random permutation, DEL_A, DEL_B, DEL_C and DEL_D. Matrices A and DEL_A have dimensions NS by NS; B and DEL_B are NS by NC; C and DEL_C are NO by NS; and D and DEL_D are NO by NC.

The primary output file for DESIGN contains an echo of the input file, and the parity check vectors and related information: the number of the parity check (i=1,...,N_PAR_CHECK), the order (NP), the eigenvalue (λ_i), and the vector itself ($\underline{w}_i$). In addition, there is a secondary output file which provides some extra information. The secondary output file includes an echo of the input file, the trace of the C_0 matrix (for each NP), a neater version of the parity vectors but with less precision, and, for each parity vector, the standard deviation, the sums of the elements which will be used as coefficients of each sensor output, and the signal-noise ratios.

C.2 PROGRAM PARCEV

Program PARCEV evaluates a set of parity check vectors, such as those generated by program DESIGN. PARCEV begins by forming a matrix containing the parity check vectors being evaluated, as follows:

$$W = \begin{bmatrix} \underline{w}_1^T \\ \underline{w}_2^T \\ \cdot \\ \cdot \\ \cdot \\ \underline{w}_{N\_PAR\_CHECK}^T \end{bmatrix} \qquad \text{(C-12)}$$

Next, the Auto-Correlation Function Matrix (ACF) is formed, as in DESIGN (Eq. C-8), and the covariance matrix of the parity checks is computed:

$$C_\nu = W \cdot ACF \cdot W^T \qquad \text{(C-13)}$$

Then, for each parity check vector $\underline{w}_i$, the sum of the coefficients of each sensor output j are computed:

$$c_{ij} = \underline{w}_{i_j} + \underline{w}_{i_{NO+j}} + \cdots + \underline{w}_{i_{NP.NO+j}}, \quad \text{for } i = 1,\ldots,N\_PAR\ CHECK, \quad j = 1,\ldots,NO \qquad \text{(C-14a)}$$

and from these coefficients, the expected, or average, residual vector is formed, for each possible sensor bias failure, positive, negative or zero (no failure), as follows:

$$\bar{\nu}_j = BFM_j \begin{bmatrix} c_{i,1} \\ c_{i,2} \\ \cdot \\ \cdot \\ \cdot \\ c_{i,NO} \end{bmatrix}, \quad \text{for } j = -NO,\ldots,-1,0,1,\ldots,NO \qquad \text{(C-14b)}$$

where:

$$BFM_{(-j)} = -BFM_j \tag{C-14c}$$

$$BFM_0 = 0 \tag{C-14d}$$

For any pair of average residual vectors, the Bhattacharyya Distance, which measures the distance between two multivariate probability distributions of specified mean and covariance, can be computed:

$$B_{ij} = \frac{1}{8} (\bar{\nu}_i - \bar{\nu}_j)^T \left[\frac{1}{2} (C_{\nu i} + C_{\nu j})\right]^{-1} (\bar{\nu}_i - \bar{\nu}_j)$$

$$+ \frac{1}{2} \ln \left(\frac{\det\left[\frac{1}{2} (C_{\nu_i} + C_{\nu_j})\right]}{\sqrt{\det(C_{\nu_i}) \cdot \det(C_{\nu_j})}} \right) \tag{C-15a}$$

(Note that, in this case, $C_{\nu_i} = C_{\nu_j} = C_\nu$, so the second term is always zero.) Furthermore, the Bhattacharyya correlation between two sensor failure hypotheses, and lower and upper bounds on the probability of error in distinguishing between two hypotheses are computed as follows:

$$\rho_{ij} = e^{-B_{ij}} \tag{C-15b}$$

$$LB_{ij} = \frac{1}{2} \left(1 - \sqrt{1 - \rho_{ij}^2}\right) \tag{C-15c}$$

$$UB_{ij} = \frac{\rho_{ij}}{2} \tag{C-15d}$$

The structure of subroutine calls in program PARCEV is shown in Fig. C-3; suboutines are described in Table C-1. The input file has precisely the same format as the primary output file for DESIGN, so that it may be run immediately afterwards, if desired. The output file for program PARCEV contains an

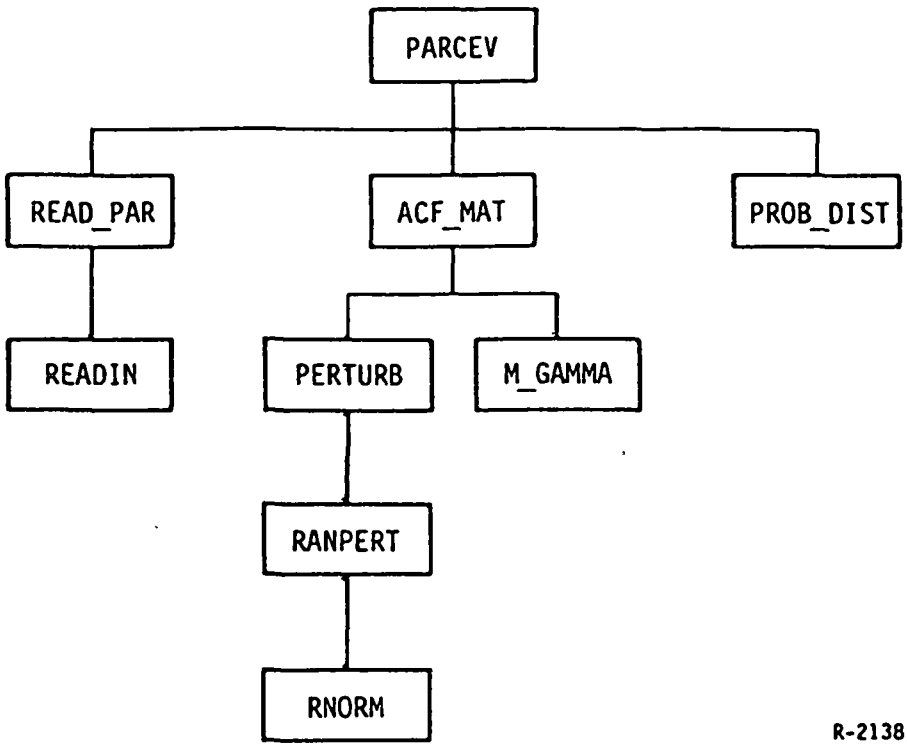


Figure C-3. Subroutine Calls Made by Program PARCEV, Excluding Library Routines

echo of the input file (i.e., the DESIGN output file), and the Bhattacharyya Distances and correlations and lower and upper bounds on the probability of error, for each pair of sensor bias failure hypotheses.

C.3 PROGRAM KFRES

Program KFRES evaluates the ability of a Kalman filter to distinguish sensor failures in much the same way that PARCEV evaluates a set of parity vectors for their ability to distinguish failures. It begins by computing, for each set of perturbed system matrices A_ℓ, B_ℓ, C_ℓ and D_ℓ (for ℓ=1,...,NL), a set of composite matrices, formed from the perturbed and unperturbed matrices as follows:

$$\overline{D}_\ell = [D_\ell - D] \tag{C-16a}$$

$$\overline{F}_\ell = \begin{bmatrix} A_\ell & 0 \\ AKC_\ell & A[I-KC] \end{bmatrix} \tag{C-16b}$$

$$\overline{G}_\ell = \begin{bmatrix} B_\ell \\ B + AK[D_\ell - D] \end{bmatrix} \tag{C-16c}$$

$$\overline{H}_\ell = \begin{bmatrix} C_\ell & -C \end{bmatrix} \tag{C-16d}$$

$$\overline{K} = \begin{bmatrix} 0 \\ AK \end{bmatrix} \tag{C-16e}$$

where K is the Kalman gain matrix.

Also, for each perturbation ℓ, the base point errors are computed:

$$x_{B_j}^{\ell} = N(0,1) \cdot \frac{Pe}{100} \cdot x_{B_j} \quad , \text{ for } j = 1,\ldots,NS \tag{C-17a}$$

$$u_{B_j}^{\ell} = N(0,1) \cdot \frac{Pe}{100} \cdot u_{B_j} \quad , \text{ for } j = 1,\ldots,NC \tag{C-17b}$$

$$y_{B_j}^{\ell} = N(0,1) \cdot \frac{Pe}{100} \cdot y_{B_j} \quad , \text{ for } j = 1,\ldots,NO \tag{C-17c}$$

where $\underline{x}_B$, $\underline{u}_B$ and $\underline{y}_B$ are the base point vectors and p_e is the percent error for base point perturbation. From the base point errors, the base point deviation vectors are computed:

$$\delta x_B{}^{\ell} = [I - A_p]\ \underline{x}_B{}^{\ell} - B_{\ell}\underline{u}_B{}^{\ell} \tag{C-18a}$$

$$\delta y_B{}^{\ell} = \underline{y}_B{}^{\ell} - C_{\ell}\underline{x}_B{}^{\ell} - D_{\ell}\underline{u}_B{}^{\ell} \tag{C-18b}$$

Next, for each possible sensor bias failure (i.e., for j = 0,1,...,NO), the augmented state vector is computed from the composite perturbed matrices and the base point deviations:

$$\underline{x}_a{}^{j\ell} = [I - \overline{F}_{\ell}]^{-1} \left[\overline{G}_{\ell}\underline{u}_B{}^{\ell} + \overline{K}\ (\underline{b}_j + \delta\underline{y}_B{}^{\ell}) + \begin{pmatrix} \delta\underline{x}_B{}^{\ell} \\ 0 \end{pmatrix} \right] \tag{C-19a}$$

where:

$$\underline{b}_j = BFM_j \cdot \underline{e}_j \tag{C-19b}$$

and $\underline{e}j$ is a unit vector in the jth direction. The augmented covariance matrix must be solved for in the following equation:

$$C_a{}^{\ell} = \overline{F}_{\ell}\ C_a{}^{\ell}\ \overline{F}_{\ell}{}^{T} + \overline{G}_{\ell}\ Q\ \overline{G}_{\ell}{}^{T} + \overline{K}\ R\ \overline{K}^{\ T} \tag{C-20}$$

From the augmented state and covariance, the Kalman filter mean (or residual) and autocovariance function are computed:

$$\underline{\nu}^{j\ell} = \overline{H}_{\ell}\ \underline{x}_a{}^{j\ell} + \overline{D}_{\ell}\ \underline{u}_B{}^{\ell} + \delta\underline{y}_B{}^{\ell} + b_j \tag{C-21a}$$

$$C_{\nu}{}^{j\ell} = \overline{H}_{\ell}\ [C_a{}^{\ell}\ (\overline{F}_{\ell}{}^{i})^{T}]\ \overline{H}_{\ell}{}^{T} + (\overline{D}_{\ell}\ Q\ \overline{D}_{\ell}{}^{T} + R)\ \delta_{io} \tag{C-21b}$$

$$+ (\overline{D}_{\ell}\ Q\ [\overline{F}_{\ell}{}^{i-1}\ \overline{G}_{\ell}]^{T}\ \overline{H}_{\ell}{}^{T} + R\ [\overline{F}_{\ell}{}^{i-1}\ \overline{K}]^{T}\ \overline{H}_{\ell}{}^{T})\ \delta_{in,n\neq 0}$$

where i is the time lag (i.e., $c_\nu{}^{j\ell} = \text{cov}\,[\nu^{i\ell}(k)\ \nu^{j\ell}(k+i)^T]$). Then, the average residual and covariance over all perturbed models may be computed for each sensor failure (j=0,1,...,NO):

$$\bar{\nu}j = \frac{1}{NL} \sum_{\ell=1}^{NL} \underline{\nu}^{j\ell} \qquad \text{(C-22a)}$$

$$C_\nu j = \frac{1}{NL} \sum_{\ell=1}^{NL} [C_\nu{}^{j\ell} + \underline{\nu}^{j\ell} (\underline{\nu}^{j\ell})^T] - \bar{\nu}j(\bar{\nu}j)^T \qquad \text{(C-22b)}$$

From these, the Bhattacharyya distances and correlations, and lower and upper bounds on the probability of error in distinguishing sensor failures, may be computed as in program PARCEV (Eq. C-15).

The structure of subroutine calls in program KFRES is shown in Fig. C-4; subroutines are described in Table C-1. The inputs to KFRES are the inputs to DESIGN, with a few additions, shown in Fig. C-5. These include MAX_LAG, the time lag used to compute the innovations covariance (i in Eq. C-21b); the random number generation seed (KF_SEED) and percent error (PER_ERR) used to compute the base point errors; the base point vectors (XB, UB, and YB); and the Kalman gain matrix (KG). The output file contains an echo of the input file, the average residual vector an covariance matrix for each possible sensor failure, and the Bhattacharyya distance and correlation and the lower and upper bounds on the probability of error in distinguishing failures, for each pair of possible sensor failures.

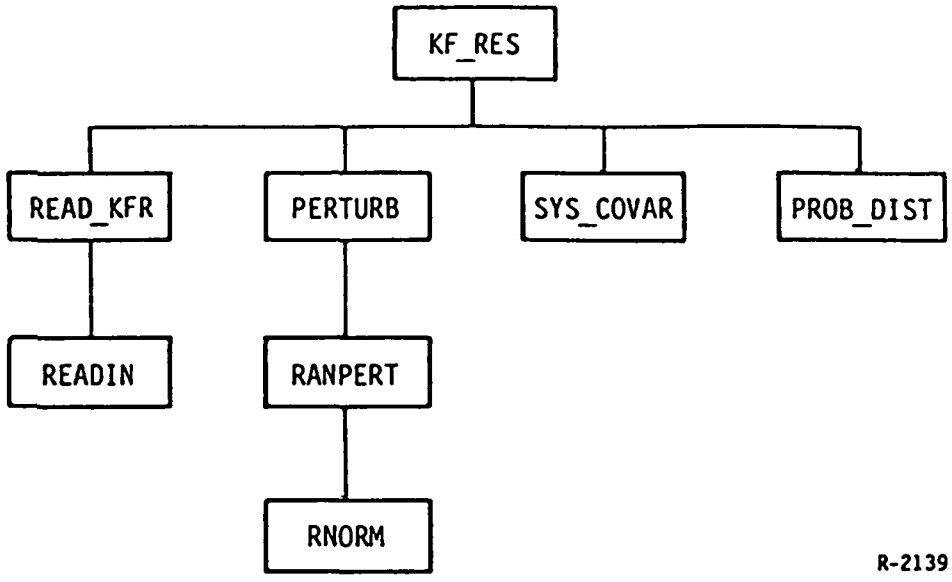


Figure C-4. Subroutine Calls Made by Program KFRES, Excluding Library Routines

```
------------------------------    KFRES INPUTS    ------------------------------

   ---   MAX_LAG   ---
    0
   ---   KF_SEED, PER_ERR   ---
    700549          0.0
   ---   XB(NS), UB(NC), YB(NO) ---
    0.8900504       0.7799951       0.1162586       8.6679444E-02
    0.5335010       0.6000000      -0.5000000       0.6000000        0.00000000E+00
    0.8900504       0.7799951       0.5165927       0.3274180       -0.8392528
   ---   KG (NS,NO)   ---
    1.487D-01       2.183D-01       7.458D-02       1.428D-01       -2.716D-01
    7.242D-02       1.061D-01       3.635D-02       6.955D-02       -1.323D-02
    4.811D-03       7.052D-03       2.407D-03       4.614D-03       -8.793D-03
    5.883D-03       8.641D-03       2.950D-03       5.643D-03       -1.075D.02
```

Figure C-5. Inputs to Program KFRES

C.4 PROGRAM SIMULATE

Program SIMULATE simulates the states, controls and outputs (sensor measurements) of the dynamic system described by Eq. C-1, including, if desired, sensor bias failures in any or all sensors, each beginning at a specified time. Two types of residuals are generated at each time step: robust parity check residuals and Kalman filter residuals. In addition, the robust parity check residuals are used to compute WSSR and SPRT statistics, in order to detect, verify and isolate sensor failures. Before robust residual generation begins, each parity vector $\underline{w}_i$ is divided into two parts: the first (NP+1)NO elements become $\underline{w}_{y_i}$, which is to be multiplied by the outputs window vector; the last (NP+1)NC elements become $\underline{w}_{u_i}$, which is to be multiplied by the controls window vector.

The dynamic system simulation used computes the deviations of the state ($\underline{\delta x}$), controls ($\underline{\delta u}$) and outputs ($\underline{\delta y}$) from a base point ($\underline{x}_b$, $\underline{u}_b$, $\underline{y}_b$) which is consistent with the system matrices. The control and associated steady state deviations are initialized as

$$\delta u_i(0) = \sqrt{q_i}\ N(0,1) \tag{C-23a}$$

$$\underline{\delta x}(0) = \underline{0} \tag{C-23b}$$

and the deviations are simulated as in the dynamic system equations (Eq. C-1), that is:

$$\underline{\delta x}(k+1) = A\ \underline{\delta x}(k) + B\ \underline{\delta u}(k) \tag{C-24a}$$

$$\underline{\delta y}(k) = C\ \underline{\delta x}(k) + D\ \underline{\delta u}(k) + \underline{v}(k) \tag{C-24b}$$

$$\underline{\delta u}(k) = \underline{n}\ (k) \tag{C-24c}$$

where the system matrices A, B, C and D, the white noise vectors $\underline{v}$ and $\underline{n}$, and the time increment Δt are as in Eq. 1.

During simulation, the control and output deviation vectors are stored over time in window vectors, as follows:

$$\underline{Y}_p(k) = \begin{bmatrix} \delta\underline{y}(k-NP) \\ \delta\underline{y}(k-NP+1) \\ \vdots \\ \delta\underline{y}(k) \end{bmatrix} \qquad \text{(C-25a)}$$

$$\underline{U}_p(k) = \begin{bmatrix} \delta\underline{u}(k-NP) \\ \delta\underline{u}(k-NP+1) \\ \vdots \\ \delta\underline{u}(k) \end{bmatrix} \qquad \text{(C-25b)}$$

The robust residuals may be computed at any time step from the parity vectors and the window vectors as

$$\nu_i(k) = \underline{w}_{y_i}^T \underline{Y}_p(k) + \underline{w}_{u_i}^T \underline{U}_p(k) \quad . \qquad \text{(C-26)}$$

Under ordinary conditions, these residuals should be close to zero; large residuals are indicative of sensor output failures or large modeling errors. These residuals are then used to compute statistics to detect and identify sensor failures, as described in the FDI equations of Appendix B.

The Kalman filter residuals are computed by a simple Kalman filter, using the unperturbed system matrices, as follows:

$$\underline{\nu}(k) = \delta\underline{y}(k) - \delta\bar{\underline{y}}(k) \qquad \text{(C-26a)}$$

where

$$\delta\bar{\underline{y}}(k) = C\ \delta\bar{\underline{x}}(k) + D\ \delta\underline{u}(k) \qquad \text{(C-26b)}$$

$$\delta\bar{\underline{x}}(k) = A\ \delta\hat{\underline{x}}(k-1) + B\,\delta\underline{u}(k-1) \qquad \text{(C-26c)}$$

$$\delta\hat{\underline{x}}(k) = \delta\ \bar{\underline{x}}(k) + K\ \underline{\nu}(k) \qquad \text{(C-26d)}$$

$$\delta\hat{\underline{x}}(0) = 0 \qquad \text{(C-26e)}$$

and K is the Kalman gain matrix.

The structure of subroutine calls in program SIMULATE is given in Fig. C-6; Table C-1 gives a brief description of the purpose of each subroutine and function. The input and output files for SIMULATE are described below.

The input file for SIMULATE has the same format as the primary output file for program DESIGN, with some additions, as shown in Fig. C-7. The first of these is TMAX, the maximum simulation time (simulation begins at time zero). The second input is PERTURB_FLAG, a logical flag (true or false) which indicates whether or not to perturb the system matrices used in Eq. C-24. Next is SIM_SEED, the initial integer value of the random number generator seed, for computing Gaussian noise.

The next inputs specify sensor bias failures during simulation. First, N_FAILURE is read in, which is the number of bias failures to occur. Then, for each failure, the following are read in: FAIL_TIME, FAIL_BIAS and FAIL_SENSOR, which are, respectively, the time that the bias occurs, the amount of the bias, and the number of the sensor in which the bias occurs. Ordinarily, a bias should cause some of the residuals to become large,

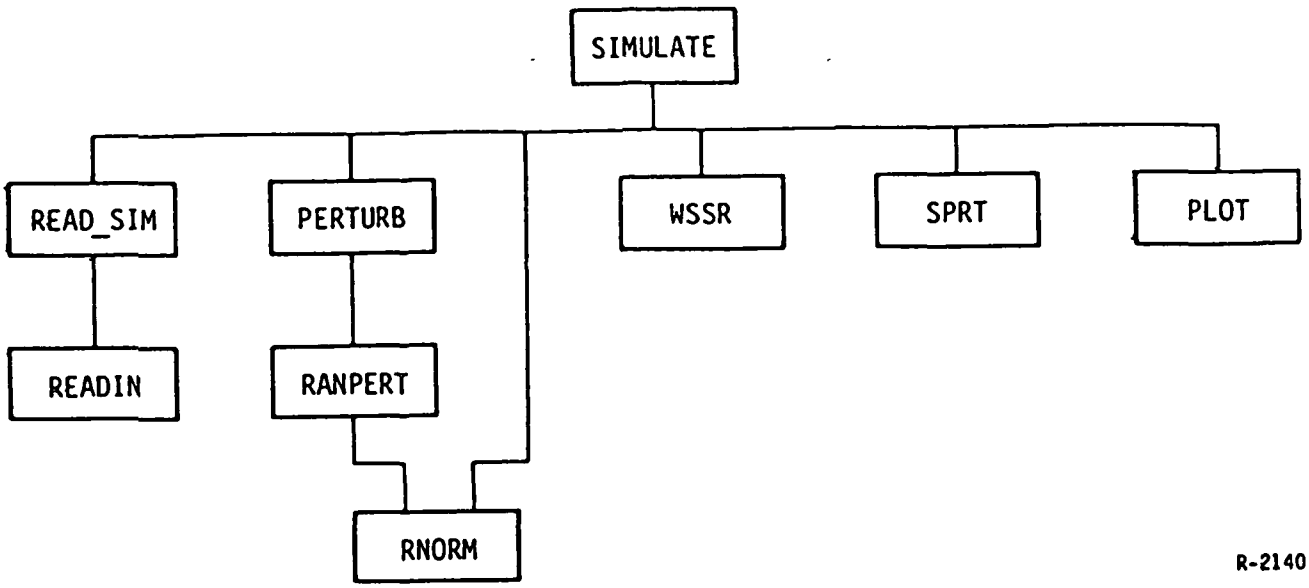


Figure C-6. Subroutine Calls Made by Program SIMULATE, Excluding Library Routines

beginning at the bias time. Following the bias failure inputs are the plotting inputs. NR_PLOT gives the number of robust parity check residuals whose simulated values should be stored at each time step in a data file, for later plotting; PLOT_INDEX specifies which residuals (NR_PLOT of them) should be stored in plot files; and KF_PLOT is a flag indicating whether or not to create plot files containing the Kalman filter residuals. Next is KFG, the Kalman gain matrix used in the Kalman filter (Eq. C-26).

The final inputs to SIMULATE are used to compute the FDI statistics. First, for each WSSR trigger statistic (N_WSSR of them), several quantities are specified. TAU_WSSR is the time constant used in the low-pass filter for the statistic; THR_WSSR is the threshold for the statistic, above which

```
----------------------   'SIMULATE' INPUTS   ------------------------
   ---   TMAX    ---
    2.0
   ---   PERTURB_FLAG   ---
    T
   ---   SIM_SEED    ---
    398219
   ---   N_FAILURE   ---
    1
   ---   FAIL_TIME, FAIL_BIAS, FAIL_SENSOR   ---
    1.0
    5.0E-3
    1
   ---   NR_PLOT, PLOT_INDEX(NR_PLOT)   ---
    5
    1    2    3    4    5
   ---   KF_PLOT   ---
    F
   ---   KFG (NS,NO)   ---
    6.725D-02   6.448D-02   2.096D-02   6.472D-02  -1.146D-01
    2.139D-02   2.981D-02   1.016D-02   2.067D-02  -3.862D-02
    1.007D-03   4.904D-04   1.133D-04   8.662D-04  -1.704D-03
    2.130D-04  -5.720D-04  -2.578D-04   5.160D-05  -2.959D-04
--------------------   WSSR INPUTS   --------------------
  ---   N_WSSR   ---
   5
  ----------   WSSR # 1  (L=1)   ----------
    ---   TAU_WSSR (L)   ---
     0.02
    ---   THR_WSSR (L)   ---
     4.0
    ---   NWS_RES(L), WS_RESN(NWS_RES,L)   ---
     1
     1
    ---   WS_COV (NWS_RES,NWS_RES,L)   ---
     1.7877E-06
  ----------   WSSR # 2  (L=2)   ----------

      .
      .
      .

--------------------   SFRT INPUTS   --------------------
  ---   N_VERIF, N_ISOL  (N_SFRT = N_VERIF+N_ISOL)   ---
   5    10
  ----------   VERIFY SFRT # 1  (L=1)   ----------
    ---   THR_SFRT (L)   ---
     20.0
    ---   NSF_RES(L), SF_RESN(NSF_RES,L)   ---
     1
     1
    ---   SF_SIG1 (NSF_RES,L)   ---
     3.4391E-03
    ---   SF_SIG2 (NSF_RES,L)   ---
     0.0000E+00
    ---   SF_COV (NSF_RES,NSF_RES,L)   ---
     1.7877E-06
  ----------   VERIFY SFRT # 2  (L=2)   ----------
```

Figure C-7. Inputs to Program SIMULATE

is indicated a possible failure; NWS_RES is the number of parity check residuals used in the statistic and WS_RESN specifies which ones; and WS_COV is the covariance matrix for the residuals used. Then, for each SPRT verify statistic (N_VERIF of them) and each SPRT isolate statistic (N_ISOL of them), the following are specified: THR_SPRT, NSP_RES, SP_RESN and SP_COV are analogous to their WSSR counterparts; SP_SIG1 and SP_SIG2 are the expected values (signatures) of the residuals used, under two different failure hypotheses. (See Appendix B for a further explanation of the use of FDI statistics).

The output file for program SIMULATE contains an echo of the input file, and an indication of when any FDI statistic passes its threshold, and when (if ever) a sensor failure is identified. In addition, plot files are created containing the parity check residuals, Kalman filter residuals, and WSSR and SPRT statistics.

*U.S. GOVERNMENT PRINTING OFFICE 1985/646-114/20414

1. Report No.	2 Government Accession No	3 Recipient's Catalog No.
NASA CR-174797		

4. Title and Subtitle	5. Report Date
Robust Detection/Isolation/Accommodation For Sensor Failures	September 1985
	6 Performing Organization Code

7. Author(s)	8 Performing Organization Report No
Jerold L. Weiss, Krishna R. Pattipati, Alan S. Willsky, John S. Eterno, and John T. Crawford	TR-213-1
	10 Work Unit No

9 Performing Organization Name and Address	11 Contract or Grant No
Alphatech, Inc. 2 Burlington Executive Center 111 Middlesex Turnpike Burlington, Massachusetts 01803	NAS 3-24078
	13 Type of Report and Period Covered
12 Sponsoring Agency Name and Address	Contractor Report
National Aeronautics and Space Administration Washington, D.C. 20546	14 Sponsoring Agency Code
	505-34-02

15 Supplementary Notes

Final report. Project Manager, Walter C. Merrill, Instrumentation and Control Technology Office, NASA Lewis Research Center, Cleveland, Ohio 44135. Appendix A - A Statistical Distance Based Methodology for Robust Parity Generation and FDI System Evaluation (TM-137) by Krishna R. Pattipati and Alan S. Willsky, Alphatech, Inc.

16 Abstract

This report presents the results of a one year study to: (1) develop a theory for Robust Failure Detection and Identification (FDI) in the presence of model uncertainty, (2) develop a design methodology which utilizes the robust FDI theory, (3) apply the methodology to a sensor FDI problem for the F-100 jet engine, and (4) demonstrate the application of the theory to the evaluation of alternative FDI schemes. Theoretical results in statistical discrimination are used to evaluate the robustness of residual signals (or parity relations) in terms of their usefulness for FDI. Furthermore, optimally robust parity relations are derived through the optimization of robustness metrics. The result is viewed as decentralization of the FDI process. A general structure for decentralized FDI is proposed and robustness metrics are used for determining various parameters of the algorithm.

17 Key Words (Suggested by Author(s))	18 Distribution Statement
FDI; Robust; Failure; Detection; DIA; Redundancy; Parity	Unclassified - unlimited STAR Category 33

19 Security Classif (of this report)	20 Security Classif (of this page)	21 No of pages	22. Price*
Unclassified	Unclassified	220	A10

*For sale by the National Technical Information Service, Springfield, Virginia 22161